



GRADO EN INGENIERÍA EN TECNOLOGÍAS INDUSTRIALES

TRABAJO DE FIN DE GRADO

**ANÁLISIS DE ESTRATEGIAS DE INVERSIÓN BASADO EN
APRENDIZAJE NO SUPERVISADO**

Autor: Marta Martínez Esteban
Directores: Juan Luis Zamora Macho
Juan Martínez Olondo

Madrid
Julio de 2019

AUTORIZACIÓN PARA LA DIGITALIZACIÓN, DEPÓSITO Y DIVULGACIÓN EN RED DE PROYECTOS FIN DE GRADO, FIN DE MÁSTER, TESINAS O MEMORIAS DE BACHILLERATO

1ª. Declaración de la autoría y acreditación de la misma.

La autora Dña. **Marta Martínez Esteban** DECLARA ser el titular de los derechos de propiedad intelectual de la obra: **ANÁLISIS DE ESTRATEGIAS DE INVERSIÓN BASADO EN APRENDIZAJE NO SUPERVISADO**, que ésta es una obra original, y que ostenta la condición de autora en el sentido que otorga la Ley de Propiedad Intelectual.

2ª. Objeto y fines de la cesión.

Con el fin de dar la máxima difusión a la obra citada a través del Repositorio institucional de la Universidad, el autor **CEDE** a la Universidad Pontificia Comillas, de forma gratuita y no exclusiva, por el máximo plazo legal y con ámbito universal, los derechos de digitalización, de archivo, de reproducción, de distribución y de comunicación pública, incluido el derecho de puesta a disposición electrónica, tal y como se describen en la Ley de Propiedad Intelectual. El derecho de transformación se cede a los únicos efectos de lo dispuesto en la letra a) del apartado siguiente.

3ª. Condiciones de la cesión y acceso

Sin perjuicio de la titularidad de la obra, que sigue correspondiendo a su autor, la cesión de derechos contemplada en esta licencia habilita para:

- a) Transformarla con el fin de adaptarla a cualquier tecnología que permita incorporarla a internet y hacerla accesible; incorporar metadatos para realizar el registro de la obra e incorporar “marcas de agua” o cualquier otro sistema de seguridad o de protección.
- b) Reproducirla en un soporte digital para su incorporación a una base de datos electrónica, incluyendo el derecho de reproducir y almacenar la obra en servidores, a los efectos de garantizar su seguridad, conservación y preservar el formato.
- c) Comunicarla, por defecto, a través de un archivo institucional abierto, accesible de modo libre y gratuito a través de internet.
- d) Cualquier otra forma de acceso (restringido, embargado, cerrado) deberá solicitarse expresamente y obedecer a causas justificadas.
- e) Asignar por defecto a estos trabajos una licencia Creative Commons.
- f) Asignar por defecto a estos trabajos un HANDLE (URL *persistente*).

4ª. Derechos del autor.

El autor, en tanto que titular de una obra tiene derecho a:

- a) Que la Universidad identifique claramente su nombre como autor de la misma
- b) Comunicar y dar publicidad a la obra en la versión que ceda y en otras posteriores a través de cualquier medio.
- c) Solicitar la retirada de la obra del repositorio por causa justificada.
- d) Recibir notificación fehaciente de cualquier reclamación que puedan formular terceras personas en relación con la obra y, en particular, de reclamaciones relativas a los derechos de propiedad intelectual sobre ella.

5ª. Deberes del autor.

El autor se compromete a:

- a) Garantizar que el compromiso que adquiere mediante el presente escrito no infringe ningún derecho de terceros, ya sean de propiedad industrial, intelectual o cualquier otro.
- b) Garantizar que el contenido de las obras no atenta contra los derechos al honor, a la intimidad y a la imagen de terceros.

- c) Asumir toda reclamación o responsabilidad, incluyendo las indemnizaciones por daños, que pudieran ejercitarse contra la Universidad por terceros que vieran infringidos sus derechos e intereses a causa de la cesión.
- d) Asumir la responsabilidad en el caso de que las instituciones fueran condenadas por infracción de derechos derivada de las obras objeto de la cesión.

6º. Fines y funcionamiento del Repositorio Institucional.

La obra se pondrá a disposición de los usuarios para que hagan de ella un uso justo y respetuoso con los derechos del autor, según lo permitido por la legislación aplicable, y con fines de estudio, investigación, o cualquier otro fin lícito. Con dicha finalidad, la Universidad asume los siguientes deberes y se reserva las siguientes facultades:

- La Universidad informará a los usuarios del archivo sobre los usos permitidos, y no garantiza ni asume responsabilidad alguna por otras formas en que los usuarios hagan un uso posterior de las obras no conforme con la legislación vigente. El uso posterior, más allá de la copia privada, requerirá que se cite la fuente y se reconozca la autoría, que no se obtenga beneficio comercial, y que no se realicen obras derivadas.
- La Universidad no revisará el contenido de las obras, que en todo caso permanecerá bajo la responsabilidad exclusiva del autor y no estará obligada a ejercitar acciones legales en nombre del autor en el supuesto de infracciones a derechos de propiedad intelectual derivados del depósito y archivo de las obras. El autor renuncia a cualquier reclamación frente a la Universidad por las formas no ajustadas a la legislación vigente en que los usuarios hagan uso de las obras.
- La Universidad adoptará las medidas necesarias para la preservación de la obra en un futuro.
- La Universidad se reserva la facultad de retirar la obra, previa notificación al autor, en supuestos suficientemente justificados, o en caso de reclamaciones de terceros.

Madrid, a 12 de julio de 2019

ACEPTA



Fdo.....

Motivos para solicitar el acceso restringido, cerrado o embargado del trabajo en el Repositorio Institucional:

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título **ANÁLISIS DE ESTRATEGIAS DE INVERSIÓN BASADO EN APRENDIZAJE NO SUPERVISADO** en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el curso académico **2018/2019** es de mi autoría, original e inédito y no ha sido presentado con anterioridad a otros efectos. El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido tomada de otros documentos está debidamente referenciada.

Fdo.:



Fecha: **12/07/ 2019**

Autorizada la entrega del proyecto
EL DIRECTOR DEL PROYECTO

Fdo.:



Fecha: 17 / 7 / 2019



GRADO EN INGENIERÍA EN TECNOLOGÍAS INDUSTRIALES

TRABAJO DE FIN DE GRADO

**ANÁLISIS DE ESTRATEGIAS DE INVERSIÓN BASADO EN
APRENDIZAJE NO SUPERVISADO**

Autor: Marta Martínez Esteban
Directores: Juan Luis Zamora Macho
Juan Martínez Olondo

Madrid
Julio de 2019

ANÁLISIS DE ESTRATEGIAS DE INVERSIÓN BASADO EN APRENDIZAJE NO SUPERVISADO

Autor: Martínez Esteban, Marta.

Directores: Zamora Macho, Juan Luis.

Martínez Olondo, Juan.

Entidad colaboradora: ICAI- Universidad Pontificia Comillas.

RESUMEN DEL PROYECTO

Introducción

Este proyecto consiste en el análisis de estrategias de inversión basado en aprendizaje no supervisado.

Motivación

Una de las motivaciones de este proyecto es detectar inversiones rentables a partir de la evolución del precio de un gran volumen de acciones pertenecientes a diferentes mercados internacionales. Además, se profundiza en el conocimiento de las técnicas de aprendizaje automático.

Objetivos

Los objetivos del proyecto son:

- Programar e integrar en una aplicación de Matlab los modelos de predicción basados en aprendizaje no supervisado: K-means y GMM.
- Desarrollar una estrategia de inversión basada en la predicción de estos modelos.
- Optimizar la rentabilidad conjunta del modelo de predicción y la estrategia de inversión

Para ello se dispone de 1300 acciones procedentes de 17 mercados internacionales y de las que se dispone de precio de apertura, cierre, volumen, moneda, etc.

Estado del arte

El aprendizaje automático consiste en diseñar algoritmos que encuentren una tendencia natural o modelo de comportamiento en los datos de un proceso para predecir su evolución futura o tomar mejores decisiones que conduzcan a su optimización.

Existen dos tipos de aprendizaje automático: aprendizaje supervisado y aprendizaje no supervisado.

El aprendizaje supervisado consiste en desarrollar modelos de predicción que relacionan entradas y salidas. Así se conseguirá predecir las salidas en el futuro a partir de los datos de las entradas disponibles.

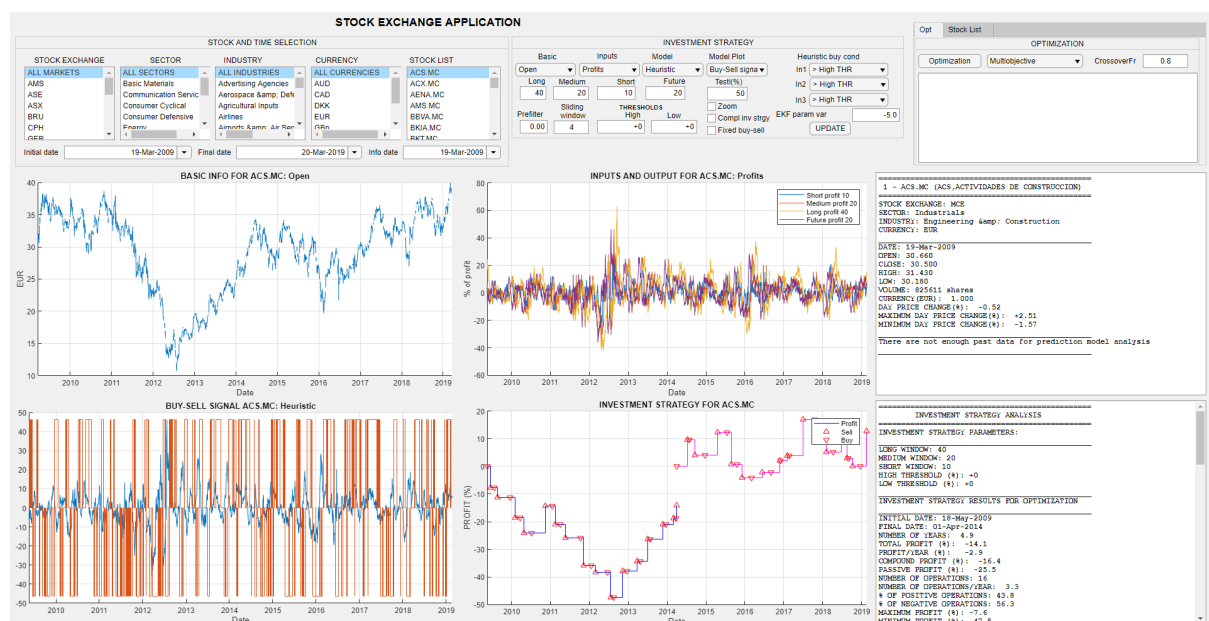
Por otro lado, en el aprendizaje no supervisado, y que se desarrolla en el proyecto, el sistema no dispone previamente de criterios con los que clasificar los datos. Se necesita agrupar los datos de entrada para poder predecir su comportamiento futuro. Una técnica de aprendizaje no supervisado es el clustering el cual trata de organizar en grupos, o clusters, los datos de entrada que se parecen de alguna manera. Dentro del clustering existen diferentes técnicas de agrupación como K-means, Gaussian Mixture Models (GMM) o redes neuronales.

Otras de las aplicaciones de aprendizaje no supervisado al ámbito financiero es el price clustering, entre otros. Este trata de predecir la tendencia natural de los precios de las acciones a agruparse en número redondos.

Metodología

Descripción de la aplicación

Como ya se ha comentado, se van a implementar estos métodos en una aplicación de Matlab cuyo objetivo consiste en definir, analizar y optimizar estrategias de inversión sobre una acción concreta.



Interfaz de la aplicación utilizada

En la parte superior izquierda se puede seleccionar la acción con la que se quiere trabajar según el mercado, el sector, la industria o la moneda.

En la parte superior central se seleccionan las entradas del modelo y el modelo de predicción con el que se quiere trabajar.

La ventana superior izquierda permite la optimización de metaparámetros.

Cada una de las gráficas representan el modelo de predicción y la estrategia de inversión correspondiente.

En la parte derecha de la aplicación hay dos partes. La primera son los datos por pantalla de la predicción de un día concreto mientras que la segunda proporciona información sobre la estrategia de inversión.

Modelo de predicción

Entradas

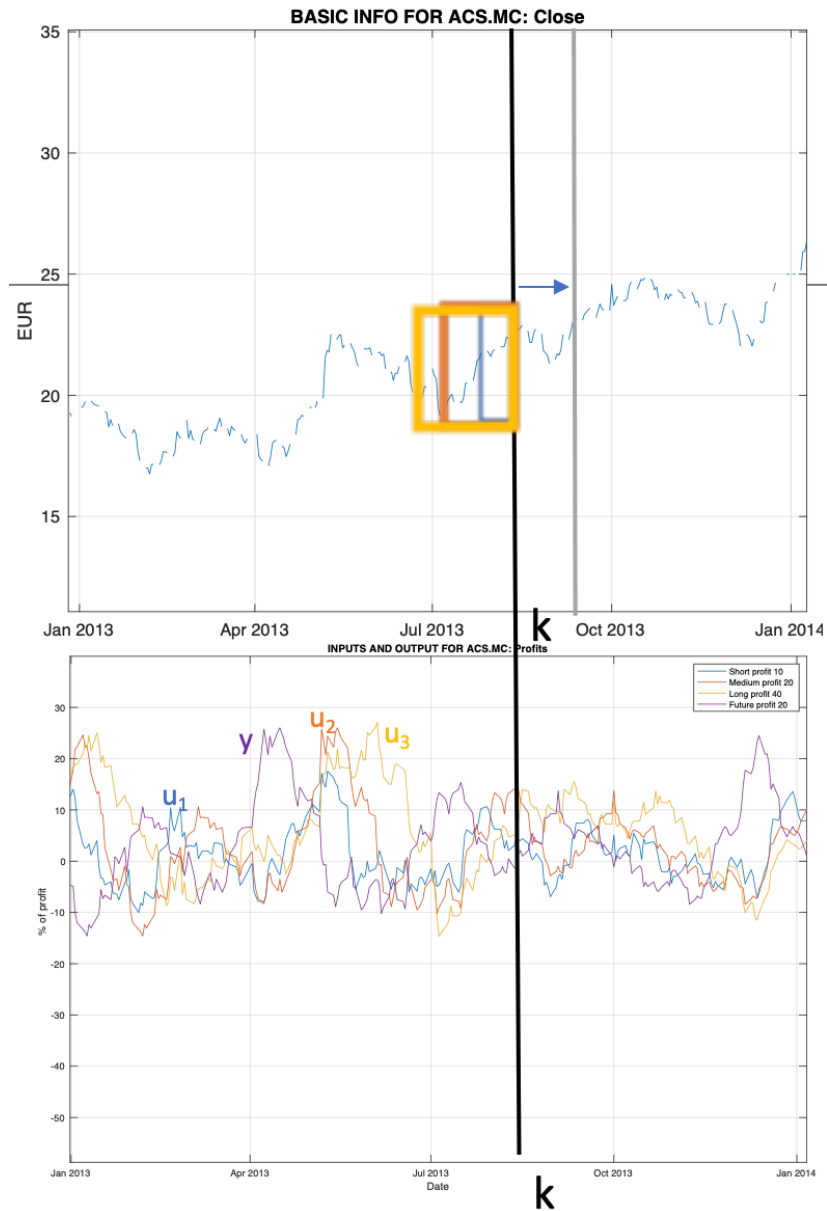
El modelo de predicción es una función matemática que depende de 3 entradas y aporta como salida la rentabilidad futura predicha.



Función matemática del modelo de predicción

Las entradas pueden expresarse de diferentes formas. Una de ellas es, por ejemplo, según las rentabilidades en tres ventanas diferentes.

Para calcular las entradas en una sesión concreta k , se utilizan los datos de precios de cierre, mostrados en la parte superior de la siguiente figura, y se calcula la rentabilidad en una ventana larga, otra media y otra corta. Esto se calcula comprando al principio de cada ventana y vendiendo al final de la misma. Así se genera la gráfica inferior que contiene las entradas de cada uno de los días y la rentabilidad real y a 20 días en este caso.

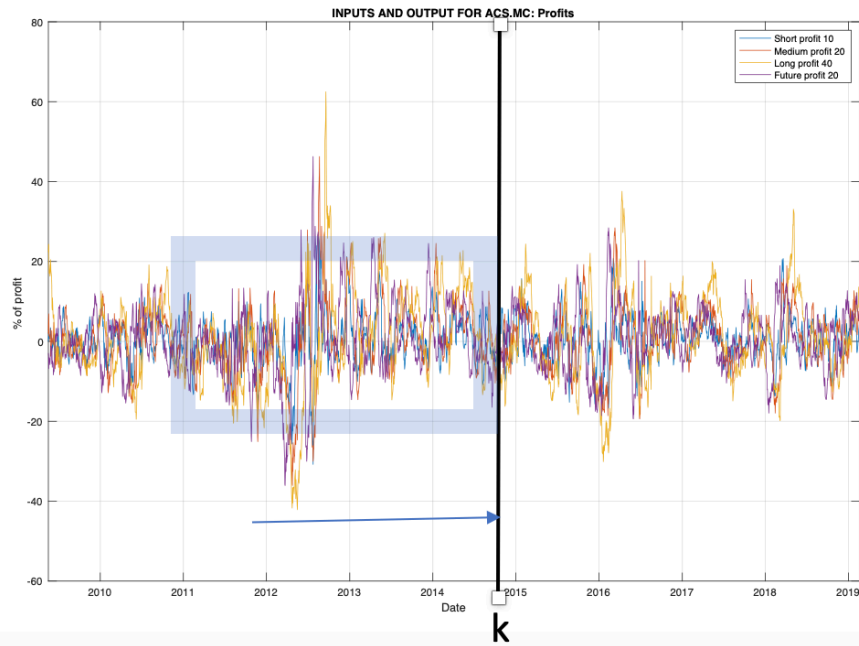


Precio de cierre y entradas y salida del modelo de predicción para ACS.MC

Clustering

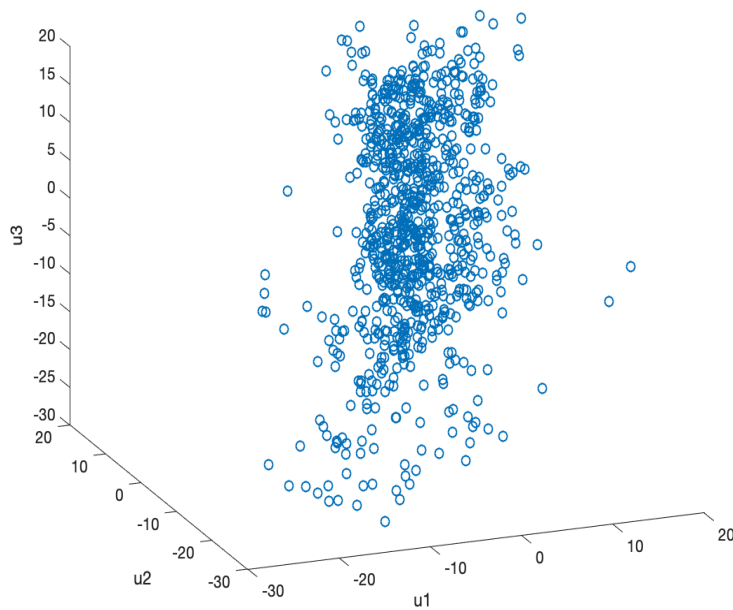
El entrenamiento de la función se realiza con una ventana de datos deslizante de 4 años, en este caso, y que permite ir adaptando el modelo en función de los cambios en la cotización de la acción. Esto queda ilustrado en la siguiente figura.

De estos 4 años se dispone de las 3 entradas definidas anteriormente y de la rentabilidad futura a 20 días.



Entradas y salidas del modelo de predicción para ACS.MC

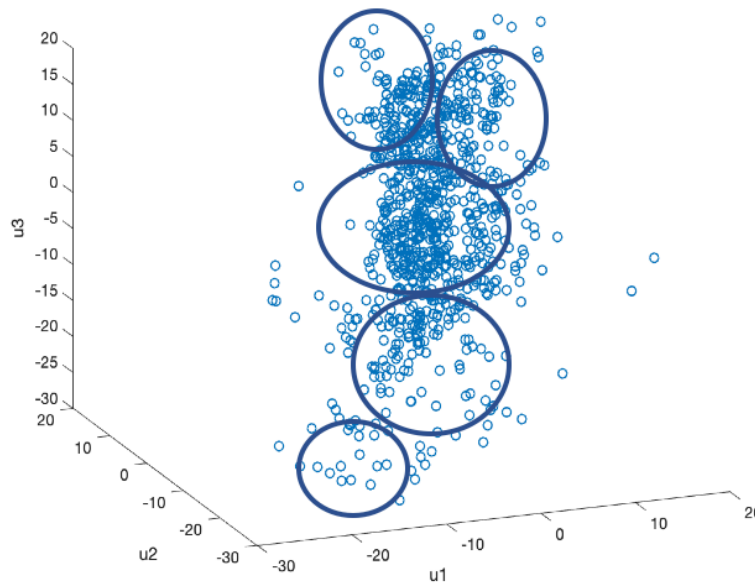
Cada día de los 4 años constituye un punto en el espacio 3D generando la siguiente gráfica.



Entradas de cada uno de los días que constituyen los 4 años para ACS.MC

Estos puntos se agrupan para formar 5 clusters siguiendo diferentes criterios según el método utilizado:

- K-means: organiza y clasifica los datos entorno a unos centroides. Esta agrupación se basa en la minimización de la distancia euclídea de los datos a sus centroides más próximos.
- GMM: utiliza un modelo de suma de distribuciones gaussianas n-dimensionales sobre los datos. Las gaussianas se ajustan en las zonas donde hay más acumulación de puntos.



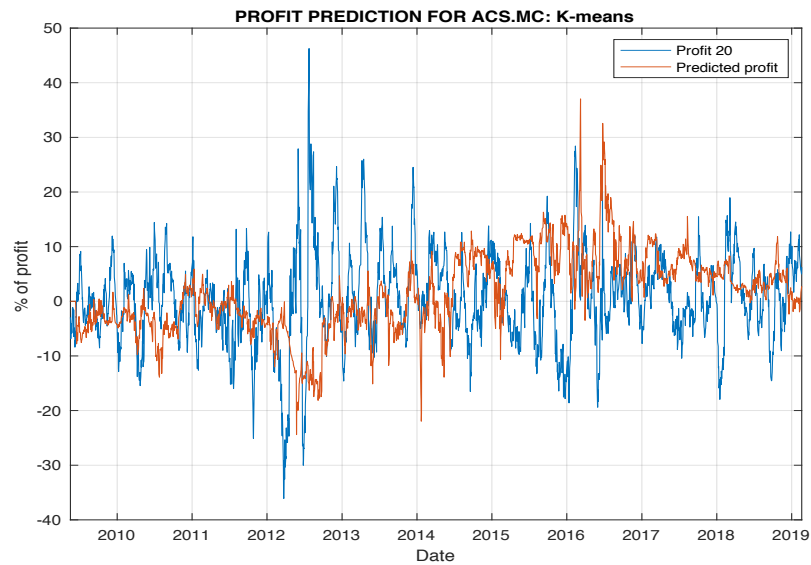
Formación de los 5 clusters para ACS.MC

Predicción

Cada uno de los clusters formados con los puntos de los 4 años tiene una rentabilidad asignada como media de las rentabilidades futuras de los puntos que pertenecen a ese cluster.

Para asignar la rentabilidad predicha a un día concreto k , se ve a qué cluster pertenecen sus coordenadas y se le asigna la rentabilidad de ese cluster como predicción futura.

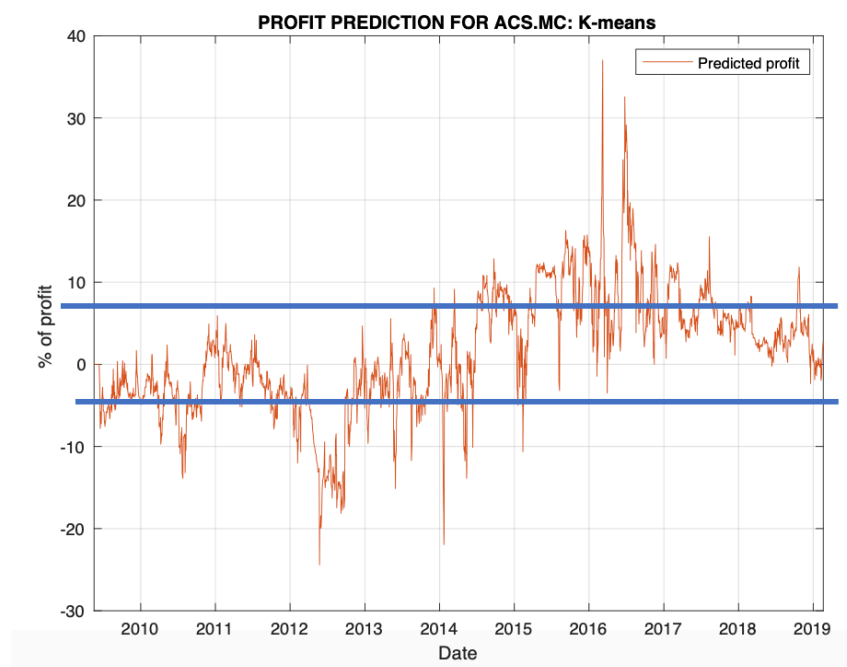
El modelo se mueve un día, los clusters se actualizan y se realiza el mismo proceso con la siguiente sesión. Así, se genera la siguiente gráfica que contiene la rentabilidad predicha como curva naranja y la rentabilidad real en azul.



Predicción de la rentabilidad futura para ACS.MC

Inversión

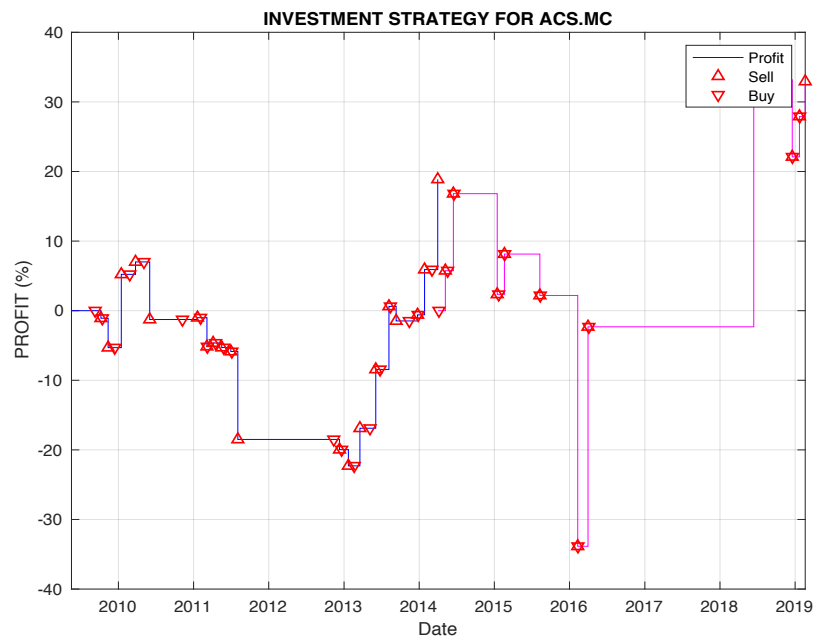
Con la curva naranja de rentabilidad predicha se calcula la señal de compra-venta. Para ello se sitúan dos umbrales y se compra cuando la rentabilidad predicha esté por encima, vendiendo cuando baje más del umbral inferior.



Señal de compra-venta para ACS.MC

Con la señal de compra-venta se genera la estrategia de inversión. Este proceso depende metaparámetros, tales como las entradas, los umbrales de inversión, etc.

Los 10 años se separan en 2 subconjuntos. En el primero de ellos se optimizan los metaparámetros y en el segundo se evalúa lo que ha pasado.



Estrategia de inversión para ACS.MC

Resultados

Se han optimizado los metaparámetros según dos métodos diferentes tanto para la técnica K-means como para GMM: optimización multiobjetivo y optimización restringida con racha negativa 10%.

La optimización multiobjetivo es una función que busca maximizar la rentabilidad y minimizar la peor racha negativa de la rentabilidad total.

Por otro lado, la optimización restringida limita la racha negativa con respecto al beneficio total. La racha negativa total en este caso tiene que ser menor del 10%.

Para optimizar los metaparámetros se han utilizado los siguientes valores: Enagás, MasterCard Incorporated, Kraft Heinz Company y UnitedHealth Group. Por otro lado, para validar se han tomado las siguientes acciones: Facebook, Visa, Abbvie Inc y Abbot Laboratories.

Los datos obtenidos tras la optimización son los siguientes:

	Optimización multiobjetivo	Optimización restringida y racha negativa 10%
Ventana larga (días)	48	40
Ventana media (días)	32	36
Ventana corta (días)	20	35
Ventana futura (días)	32	8
Prefiltro	0.18	0.26
Umbral de compra (%)	-1	-1
Umbral de venta (%)	-1	-1

Valores de los metaparámetros optimizados

En las siguientes tablas aparecen los datos de las rentabilidades y rachas negativas de las acciones mencionadas previamente tras la optimización de metaparámetros.

Se ha señalado de azul, aquellos conjuntos de racha negativa y rentabilidad que proporcionan buenos resultados debido a que el número de operaciones realizadas no es pequeño y además la racha negativa es pequeña comparado con la rentabilidad que produce.

Por lo tanto, aquellos conjuntos que no quedan seleccionados son debido a que no aportan buenos resultados: o su racha negativa es grande en relación con la rentabilidad o el número de operaciones realizadas es pequeño debido a que la racha negativa se ha restringido demasiado.

Además, para el método K-means se va a utilizar un ratio para estudiar la rentabilidad obtenida por día invertido para cada una de las acciones. El cociente se define de la siguiente forma:

$$Ratio = \frac{Rentabilidad}{Días_inversión_activa}$$

Para calcular el ratio anualizado se multiplica el ratio diario por 250 días.

	ENTRENAMIENTO				TEST			
VALORES	RENTABILIDAD (%)	RACHA NEGATIVA (%)	RATIO	RATIO ANUAL	RENTABILIDAD (%)	RACHA NEGATIVA (%)	RATIO	RATIO ANUAL
Enagás	60.6	-10.3	0.049	12.25	16.5	-14.5	0.0134	3.35
MasterCard Incorporated	322.8	0	0.265	66.25	238	0	0.196	49
The Kraft Heinz Company	20.2	0	0.0478	11.95	-53.3	-33.9	-0.205	-51.25
UnitedHealth Group Incorporated	130.5	-10.7	0.118	29.5	297.5	0	0.244	61
TOTAL	133.5			30	124.7			15.52
Facebook	359.3	0	0.468	117	52.1	0	0.064	16
Visa	164.3	0	0.141	35.25	201.4	0	0.165	41.25
Abbvie Inc	69.2	0	0.093	23.25	37.8	0	0.051	12.75
Abbot Laboratories	57.6	-0.75	0.0481	12.025	103.6	0	0.085	21.25
TOTAL	162.6			46.88	98.7			22.81

Optimización multiobjetivo para método K-means

Utilizando el ratio definido anteriormente, se observa que el conjunto de acciones utilizadas para la optimización multiobjetivo producen unos ratios anualizados de 30% y 15.52% en el entrenamiento y test respectivamente.

Si se estudian aquellas acciones utilizadas para la validación los ratios anualizados son 46.88% para el entrenamiento y 22.81% para el test.

Aunque en ambos casos difieran entre sí, estos valores anuales son muy buenos.

	ENTRENAMIENTO				TEST			
VALORES	RENTABILIDAD (%)	RACHA NEGATIVA (%)	RATIO	RATIO ANUAL	RENTABILIDAD (%)	RACHA NEGATIVA (%)	RATIO	RATIO ANUAL
Enagás	34.3	-31.9	0.03	7.5	-8.7	-19.1	- 0.0075	-1.88
Mastercard Incorporated	220.9	-15.2	0.185	46.25	292.2	0	0.237	59.25
The Kraft Heinz Company	10.1	-9.4	0.026	6.5	-61.2	-28.4	-0.217	-54.25
UnitedHealth Group Incorporated	113.8	-19.4	0.102	25.5	195.9	0	0.16	40
TOTAL	94.77			21.43	104.6			10.78
Facebook	358.6	-16.7	0.487	121.75	49.1	0	0.059	14.75
Visa	109.5	-16.4	0.095	23.75	261.7	0	0.212	53
Abbvie Inc	67.3	-3.6	0.09	22.5	34.8	-28.9	0.045	11.25
Abbot Laboratories	66.3	0	0.054	13.5	100.1	-18.5	0.083	20.75
TOTAL	150.44			45.37	111.5			24.94

Optimización restringida y con racha negativa 10% para K-means

Teniendo en cuenta el ratio para ambos casos, con las acciones utilizadas para optimizar el ratio anual es de 21.43% en el entrenamiento y 10.78% en el test y, en el caso de las acciones utilizadas para la validación estos valores son 45.37% en el entrenamiento y 24.94% en el test.

	ENTRENAMIENTO		TEST	
VALORES	RENTABILIDAD (%)	RACHA NEGATIVA (%)	RENTABILIDAD (%)	RACHA NEGATIVA (%)
Enagás	53.94	-16.34	8.28	0
Mastercard Incorporated	236.87	0	238.01	0
The Kraft Heinz Company	21.6	-2.95	-37.2	-31.73
UnitedHealth Group Incorporated	120.03	0	297.48	0
TOTAL	108.11		126.64	
Facebook	301.86	0	52.15	0
Visa	157.58	0	201.36	0
Abbvie Inc	43.66	-19.54	28.12	-10.34
Abbot Laboratories	36.39	-17.02	129.22	0
TOTAL	134.87		102.71	

Optimización multiobjetivo para GMM

En la optimización multiobjetivo para GMM los datos conjuntos del entrenamiento tanto en las acciones utilizadas en optimización como validación son mucho más positivos que los obtenidos en el test.

	ENTRENAMIENTO		TEST	
VALORES	RENTABILIDAD (%)	RACHA NEGATIVA (%)	RENTABILIDAD (%)	RACHA NEGATIVA (%)
Enagás	-15.33	-52.73	4.25	-14.6
Mastercard Incorporated	92.43	-29.34	221.81	0
The Kraft Heinz Company	6.63	-11.7	-67.75	-57.69
UnitedHealth Group Incorporated	117.49	-15.52	200.57	0
TOTAL	50.31		89.72	
Facebook	119.76	0	46.44	0
Visa	59.68	-54.73	261.69	0
Abbvie Inc	33.66	-16.79	53.89	-12.96
Abbot Laboratories	10.71	-37.06	153.87	0
TOTAL	55.95		128.97	

Optimización restringida y con racha negativa 10% para GMM

En el caso de la optimización restringida y racha negativa 10% para GMM, la mayoría de las rachas negativas de las acciones son muy altas en comparación con las rentabilidades que generan con excepción de UnitedHealth Group Incorporated en la optimización y Facebook en la validación.

Como conclusión general, este método aplicado al GMM produce peores resultados que en K-means.

Conclusiones

Tras haber programado los modelos de predicción para K-means y GMM y haber optimizado los metaparámetros para ambos métodos utilizando tanto optimización multiobjetivo como optimización restringida con racha negativa 10% se ha observado que ambos métodos funcionan muy bien para la técnica K-means mientras que para GMM no presentan buenos resultados.

La optimización multiobjetivo para el conjunto de acciones utilizadas en la optimización de metaparámetros para K-means produce un 30% de ratio anualizado en el entrenamiento y un 15.52% de ratio anualizado en el test. Esta optimización aplicada al conjunto de acciones utilizadas para validación genera unos ratios anualizados de 46.88% y 22.81% para entrenamiento y test respectivamente.

Por otro lado, la optimización restringida con racha negativa 10% para el conjunto de acciones utilizadas en la optimización de metaparámetros para K-means genera un 21.43% de ratio anualizado en el entrenamiento y un 10.78% de ratio anualizado en el test. Esta optimización aplicada al conjunto de acciones utilizadas para validación genera unos ratios anualizados de 45.37% y 24.94% para entrenamiento y test respectivamente.

Como se puede observar todos los resultados obtenidos generan ratios anualizados muy altos.

ANALYSIS OF INVESTMENT STRATEGIES BASED ON UNSUPERVISED MACHINE LEARNING

Author: Martínez Esteban, Marta.

Directors: Zamora Macho, Juan Luis.

Martínez Olondo, Juan.

Collaborating Entity: ICAI- Universidad Pontificia Comillas.

ABSTRACT

Introduction

This project analyses investment strategies based on unsupervised machine learning.

Motivation

One of the motivations of this project is to detect profitable investments based on the evolution of the price of a large volume of shares belonging to different international markets. In addition, the knowledge of automatic learning techniques is deepened.

Objectives

The objectives of the project are:

- Program and integrate in a Matlab application the prediction models based on unsupervised machine learning: K-means and GMM.
- Develop an investment strategy based on the prediction of these models.
- Optimize the profitability of the prediction model and the investment strategy.

For this purpose, 1300 stocks from 17 international markets are available, with opening price, closing price, volume, currency, etc.

State of art

Machine learning consists of designing algorithms that find a natural trend or behavioral model in the data of a process in order to predict its future evolution or take better decisions that lead to its optimization.

There are two types of machine learning: supervised machine learning and unsupervised machine learning.

Supervised machine learning consists of developing predictive models that relate inputs and outputs. In this way, you will be able to predict outputs in the future from available input data.

On the other hand, in unsupervised machine learning, which is developed in the project, the system does not previously have criteria to classify the data. It is necessary to group the input data in order to predict its future behavior. An unsupervised machine learning technique is clustering, which tries to organize in groups, or clusters, the input data that are somehow similar. Within clustering there are different grouping techniques such as K-means, Gaussian Mixture Models (GMM) or neural networks.

Other unsupervised machine learning applications in the financial field is price clustering, among others. This tries to predict the natural tendency of the prices of the shares to be grouped in round numbers.

Methodology

Description of the application

As already mentioned, these methods will be implemented in a Matlab application whose objective is to define, analyze and optimize investment strategies for a specific stock.



Application interface

In the upper left side, you can select the stock you want to work with according to market, sector, industry or currency.

In the upper central side, you select the inputs of the model and the prediction model you want to work with.

The upper left side allows the optimization of meta-parameters.

Each of the graphs represents the prediction model and the corresponding investment strategy.

On the right side of the application there are two parts. The first is the on-screen data of the prediction of a particular day while the second provides information on the investment strategy.

Prediction model

Inputs

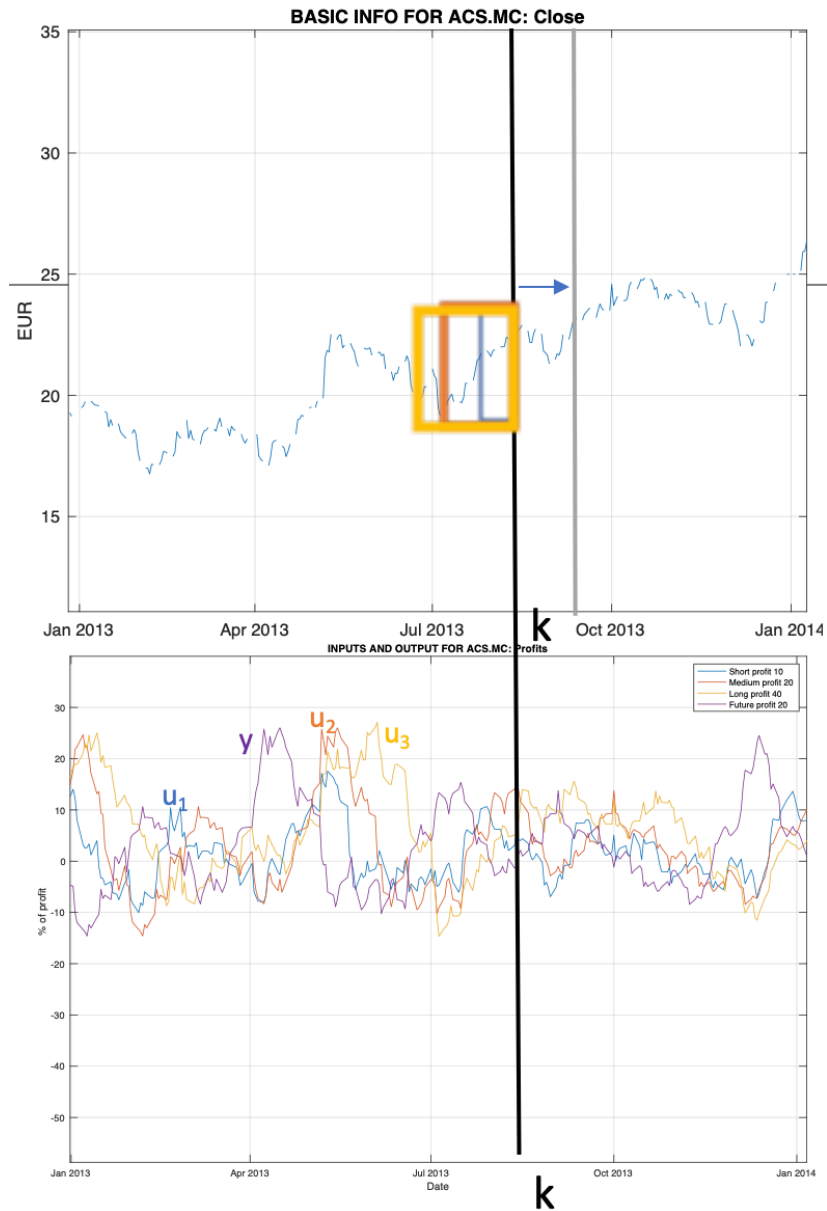
The prediction model is a mathematical function that depends on 3 inputs and provides the predicted future profitability as an output.



Mathematical function of the prediction model

Inputs can be expressed in different ways. One of them is, for example, according to the yields in three different windows.

To calculate the inputs in a specific session k , the closing price data shown in the upper part of the following figure is used, and the profitability is calculated in a long window, a medium window and a short window. This is calculated by buying at the beginning of each window and selling at the end of it. This generates the lower graph containing the inputs for each day and the actual return y at 20 days in this case.

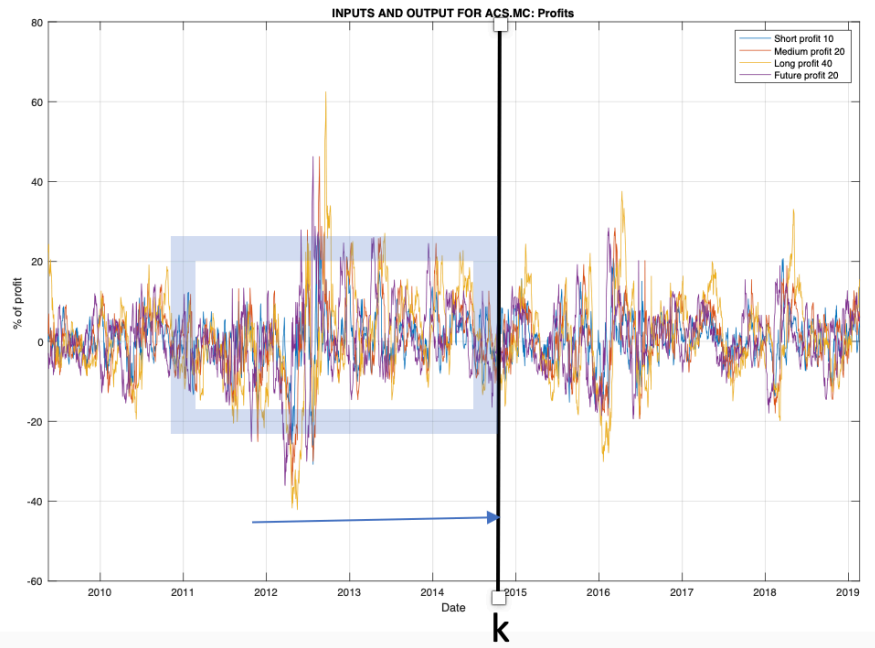


Closing price and prediction model inputs and outputs for ACS.MC

Clustering

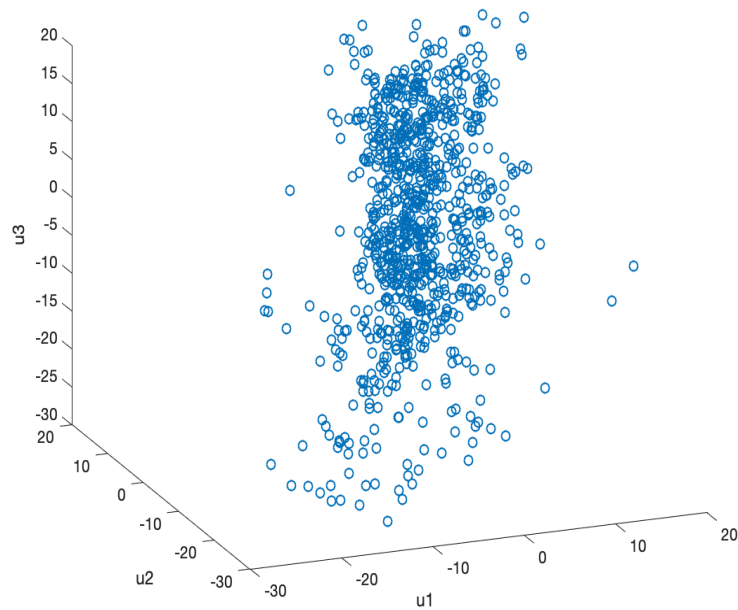
The training of the function is carried out with a sliding data window of 4 years in this case, which allows the model to be adapted according to the changes in the stock price. This is illustrated in the following figure.

Of these 4 years, the 3 inputs defined above and the future profitability after 20 days are available.



Prediction model inputs and outputs for ACS.MC

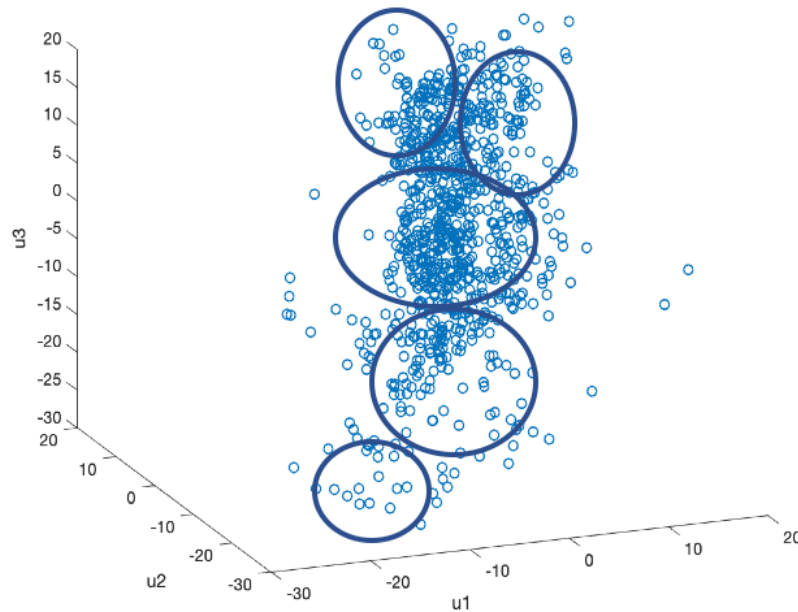
Each day of the 4 years constitutes a point in the 3D space generating the following graph.



Inputs for the days constituting the 4 years for ACS.MC

These points are grouped to form 5 clusters according to different criteria according to the method used:

- K-means: organizes and classifies the data around some centroids. This grouping is based on the minimization of the Euclidean distance of the data to their nearest centroids.
- GMM: uses a model of sum of n-dimensional Gaussian distributions on the data. The Gaussian are adjusted in the areas where there is more accumulation of points.



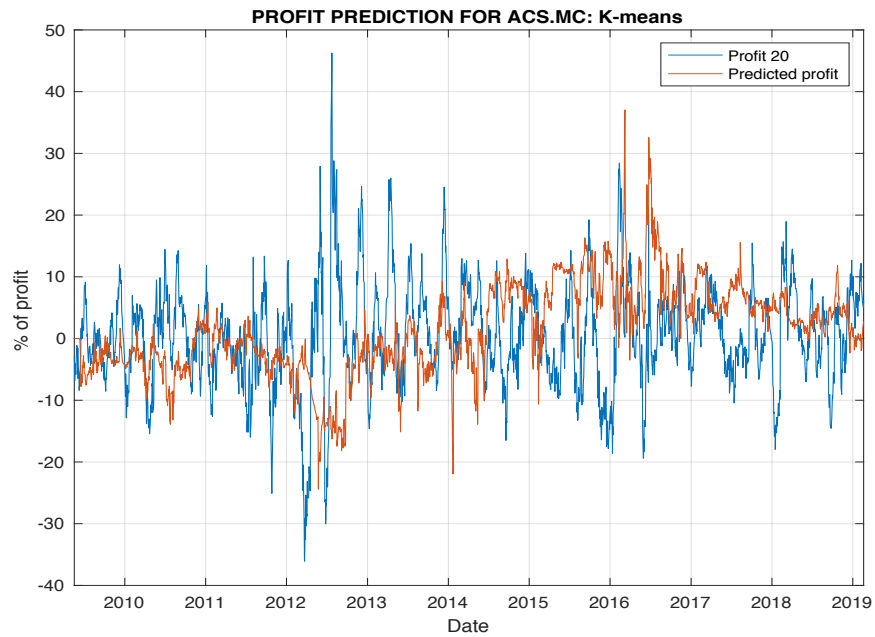
Formation of the 5 clusters for ACS.MC

Prediction

Each of the clusters formed with the points of the 4 years has a profitability assigned as an average of the future profitability of the points that belong to that cluster.

To assign the predicted profitability to a specific day k , you see which cluster its coordinates belong to and you assign the profitability of that cluster as a future prediction.

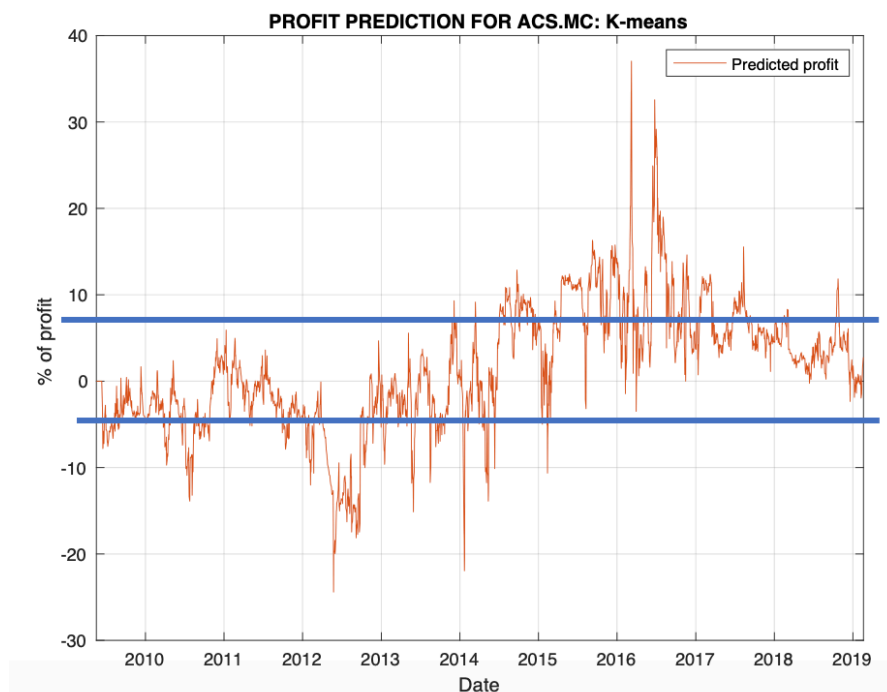
The model moves one day, the clusters are updated and the same process is carried out with the next session. Thus, the following graph is generated containing the predicted profitability as an orange curve and the real profitability as a blue curve.



Prediction of future profitability for ACS.MC

Investment

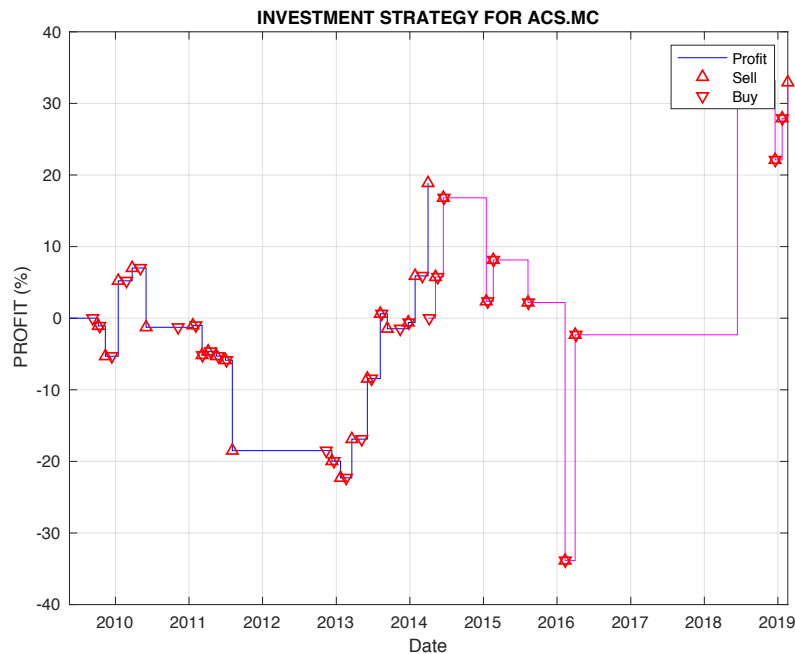
With the predicted orange yield curve, the buy-sell signal is calculated. To do this, two thresholds are placed and purchase is made when the predicted profitability is above, selling when it falls below the lower threshold.



Buy-Sell signal for ACS.MC

With the buy-sell signal, the investment strategy is generated. This process depends on meta-parameters, such as inputs, investment thresholds, etc.

The 10 years are separated into 2 subsets. In the first one the meta-parameters are optimized and in the second one what has happened is evaluated.



Investment strategy for ACS.MC

Results

The meta-parameters have been optimized according to two different methods for both the K-means technique and GMM: multi-objective optimization and restricted optimization with 10% negative streak.

Multi-objective optimization is a function that seeks to maximize profitability and minimize the worst negative profitability streak.

On the other hand, restricted optimization limits the negative streak with respect to total profit. The total negative streak in this case has to be less than 10%.

The following stocks have been used to optimize the meta-parameters: Enagás, MasterCard Incorporated, Kraft Heinz Company and UnitedHealth Group. On the other hand, the following stocks have been used to validate: Facebook, Visa, Abbvie Inc and Abbot Laboratories.

The data obtained after the optimization are as follows:

	Multi-objective optimization	Restricted optimization with 10% negative streak
Long window (days)	48	40
Medium window (days)	32	36
Short window (days)	20	35
Future window (days)	32	8
Prefilter	0.18	0.26
Umbral de compra (%)	-1	-1
Umbral de venta (%)	-1	-1

Optimized meta-parameter values

The following tables show the data on the yields and negative streaks of the previously mentioned stocks after meta-parameter optimization.

It has been pointed out in blue, those sets of negative streaks and profitability that provide good results because the number of operations carried out is not small and in addition the negative streak is small compared to the profitability it produces.

Therefore, those sets that are not selected are because they do not provide good results: either their negative streak is large in relation to profitability or the number of operations performed is small because the negative streak has been restricted too much.

In addition, for the K-means method, a ratio will be used to study the return obtained per day invested for each of the stocks. The ratio is defined as follows:

$$Ratio = \frac{Profitability}{Days_investment}$$

To calculate the annual ratio, the daily ratio is multiplied by 250 days.

	TRAINING				TEST			
STOCK LIST	PROFITABILITY (%)	NEGATIVE STREAK (%)	RATIO	ANNUAL RATIO	PROFITABILITY (%)	NEGATIVE STREAK (%)	RATIO	ANNUAL RATIO
Enagás	60.6	-10.3	0.049	12.25	16.5	-14.5	0.0134	3.35
MasterCard Incorporated	322.8	0	0.265	66.25	238	0	0.196	49
The Kraft Heinz Company	20.2	0	0.0478	11.95	-53.3	-33.9	-0.205	-51.25
UnitedHealth Group Incorporated	130.5	-10.7	0.118	29.5	297.5	0	0.244	61
TOTAL	133.5			30	124.7			15.52
Facebook	359.3	0	0.468	117	52.1	0	0.064	16
Visa	164.3	0	0.141	35.25	201.4	0	0.165	41.25
Abbvie Inc	69.2	0	0.093	23.25	37.8	0	0.051	12.75
Abbot Laboratories	57.6	-0.75	0.0481	12.025	103.6	0	0.085	21.25
TOTAL	162.6			46.88	98.7			22.81

Multi-objective optimization for K-means

Using the ratio defined above, it can be seen that the set of stocks used for multiobjective optimization produce annual ratios of 30% and 15.52% in training and test respectively.

If those stocks used for validation are studied, the annual ratios are 46.88% for training and 22.81% for testing.

Although in both cases they differ from each other, these annual values are still very good.

	TRAINING				TEST			
STOCK LIST	PROFITABILITY (%)	NEGATIVE STREAK (%)	RATIO	ANNUAL RATIO	PROFITABILITY (%)	NEGATIVE STREAK (%)	RATIO	ANNUAL RATIO
Enagás	34.3	-31.9	0.03	7.5	-8.7	-19.1	-0.0075	-1.88
Mastercard Incorporated	220.9	-15.2	0.185	46.25	292.2	0	0.237	59.25
The Kraft Heinz Company	10.1	-9.4	0.026	6.5	-61.2	-28.4	-0.217	-54.25
UnitedHealth Group Incorporated	113.8	-19.4	0.102	25.5	195.9	0	0.16	40
TOTAL	94.77			21.43	104.6			10.78
Facebook	358.6	-16.7	0.487	121.75	49.1	0	0.059	14.75
Visa	109.5	-16.4	0.095	23.75	261.7	0	0.212	53
Abbvie Inc	67.3	-3.6	0.09	22.5	34.8	-28.9	0.045	11.25
Abbot Laboratories	66.3	0	0.054	13.5	100.1	-18.5	0.083	20.75
TOTAL	150.44			45.37	111.5			24.94

Restricted optimization and negative streak 10% for K-means

Taking into account the ratio for both cases, with the stocks used to optimize the annual ratio is 21.43% in the training and 10.78% in the test and, in the case of the stocks used for validation these values are 45.37% in the training and 24.94% in the test.

	TRAINING		TEST	
STOCK LIST	PROFITABILITY (%)	NEGATIVE STREAK (%)	PROFITABILITY (%)	NEGATIVE STREAK (%)
Enagás	53.94	-16.34	8.28	0
Mastercard Incorporated	236.87	0	238.01	0
The Kraft Heinz Company	21.6	-2.95	-37.2	-31.73
UnitedHealth Group Incorporated	120.03	0	297.48	0
TOTAL	108.11		126.64	
Facebook	301.86	0	52.15	0
Visa	157.58	0	201.36	0
Abbvie Inc	43.66	-19.54	28.12	-10.34
Abbot Laboratories	36.39	-17.02	129.22	0
TOTAL	134.87		102.71	

Multi-objective optimization for GMM

In multi-objective optimization for GMM the training data sets both in the stocks used in optimization and validation are much more positive than those obtained in the test.

STOCK LIST	TRAINING		TEST	
	PROFITABILITY (%)	NEGATIVE STREAK (%)	PROFITABILITY (%)	NEGATIVE STREAK (%)
Enagás	-15.33	-52.73	4.25	-14.6
Mastercard Incorporated	92.43	-29.34	221.81	0
The Kraft Heinz Company	6.63	-11.7	-67.75	-57.69
UnitedHealth Group Incorporated	117.49	-15.52	200.57	0
TOTAL	50.31		89.72	
Facebook	119.76	0	46.44	0
Visa	59.68	-54.73	261.69	0
Abbvie Inc	33.66	-16.79	53.89	-12.96
Abbot Laboratories	10.71	-37.06	153.87	0
TOTAL	55.95		128.97	

Restricted optimization and negative streak 10% for GMM

In the case of restricted optimization and negative streak 10% for GMM, most of the negative streak of stocks are very high compared to the yields they generate with the exception of UnitedHealth Group Incorporated in optimization and Facebook in validation.

As a general conclusion, this method applied to GMM does not get better results than K-means.

Conclusions

After having programmed the prediction models for K-means and GMM and having optimized the metaparameters for both methods using both multi-objective optimization and restricted optimization with a 10% negative streak it has been observed that both methods work very well for the K-means clustering technique while for GMM they do not show good results.

The multi-objective optimization for the set of stocks used in the optimization of meta-parameters for K-means produces 30% annual ratio in training and 15.52% annual ratio in the test. This optimization applied to the set of stocks used for validation generates annual ratios of 46.88% and 22.81% for training and test respectively.

On the other hand, the restricted optimization with a negative streak of 10% for the set of stocks used in the optimization of meta-parameters for K-means generates 21.43% of annual ratio in training and 10.78% of annual ratio in the test. This optimization applied to the set of stocks used for validation generates annual ratios of 45.37% and 24.94% for training and test respectively.

As can be seen, all the results obtained generate very high annual ratios.

ÍNDICE DE CONTENIDO

1. Introducción	1
1.1. Motivación	1
1.2. Objetivos.....	2
1.3. Metodología.....	2
1.4. Recursos.....	3
1.5. Estructura de la memoria	3
2. Estado del arte.....	5
2.1. El mercado bursátil.....	5
2.1.1. Análisis fundamental	6
2.1.2. Análisis técnico	6
2.2. Tipos de aprendizaje automático	6
2.2.1. Aprendizaje supervisado	6
2.2.2. Aprendizaje no supervisado	7
3. Descripción de la aplicación	11
3.1. Entradas y salidas al modelo de predicción.....	15
3.2. Modelo de predicción.....	22
3.3. Estrategia de inversión	23
3.4. Ejemplo utilizado para la ilustración del proceso para los métodos K-means y GMM	24
4. K-means.....	27
5. GMM	41
6. Análisis de resultados	53
7. Conclusiones y futuros desarrollos.....	67
Bibliografía	69

ÍNDICE DE FIGURAS

Fig. 3.1: Interfaz de la aplicación utilizada	11
Fig. 3.2: Función de selección de las acciones.....	11
Fig. 3.3: Configuración de opciones de la estrategia de inversión	12
Fig. 3.4: Configuración de optimización	12
Fig. 3.5: Gráficas e información de la estrategia de inversión.....	12
Fig. 3.6: Precio de cierre y apertura para ACS.MC	14
Fig. 3.7: Precio de cierre y apertura normalizados para ACS.MC	14
Fig. 3.8: Rentabilidades para ACS.MC	15
Fig. 3.9: Diferencias de rentabilidades para ACS.MC	16
Fig. 3.10: Rentabilidad más reciente, máxima y mínima para ACS.MC	17
Fig. 3.11: Diferencias rentabilidad más reciente, máxima y mínima para ACS.MC.....	17
Fig. 3.12: Medias de las rentabilidades para ACS.MC	18
Fig. 3.13: Diferencias de las medias de rentabilidades para ACS.MC.....	19
Fig. 3.14: Regresión lineal para ACS.MC.....	20
Fig. 3.15: Rentabilidades muestreadas para ACS.MC.....	21
Fig. 3.16: Diferencia entre rentabilidades muestreadas para ACS.MC	21
Fig. 3.17: Función matemática del modelo de predicción	22
Fig. 3.18: Ejemplo de estrategia de inversión	24
Fig. 3.19: Salida por pantalla del precio de cierre y apertura para ACS.MC en el día 18 de marzo de 2016.....	25
Fig. 3.20: Entradas y salidas del modelo de predicción para ACS.MC	25
Fig. 3.21: Entradas de cada uno de los días que constituyen los 4 años para ACS.MC ..	26
Fig. 4.1: Coordenadas de los centroides para ACS.MC en el día 18 de marzo de 2016 ..	27
Fig. 4.2: Salida por pantalla de los datos de las señales de entrada para ACS.MC.....	30
Fig. 4.3: Salida por pantalla de la información del día 18 de marzo de 2016 para ACS.MC	30
Fig. 4.4: Predicción de la rentabilidad futura para ACS.MC.....	31
Fig. 4.5: Correlación entre rentabilidad predicha y rentabilidad real para ACS.MC.....	31
Fig. 4.6: Correlación entre rentabilidad predicha y rentabilidad real para DIA.MC	32
Fig. 4.7: Movimiento de los centroides para ACS.MC	32
Fig. 4.8: Evolución temporal de las coordenadas u_1 , u_2 , u_3 para cada uno de las clusters para ACS.MC.....	33
Fig. 4.9: Evolución temporal de las tres entradas para cada uno de los 5 clusters para ACS.MC	34
Fig. 4.10: Evolución de las rentabilidades de cada cluster para ACS.MC	35
Fig. 4.11: Evolución de las rentabilidades promediadas de cada cluster para ACS.MC ..	35
Fig. 4.12: Señal de compra-venta para ACS.MC	36
Fig. 4.13: Estrategia de inversión para ACS.MC.....	36

Fig. 4.14: Estrategia de inversión complementaria para ACS.MC	37
Fig. 4.15: Estrategia de inversión fija para ACS.MC.....	37
Fig. 4.16: Salida por pantalla de la información de la estrategia de inversión para ACS.MC.....	40
Fig. 5.1: Salida por pantalla de los datos de las señales de entrada para ACS.MC.....	42
Fig. 5.2: Salida por pantalla de la información del día 18 de marzo de 2016 para ACS.MC.....	42
Fig. 5.3: Predicción de la rentabilidad futura para ACS.MC.....	43
Fig. 5.4: Correlación entre rentabilidad predicha y rentabilidad real para ACS.MC.....	43
Fig. 5.5: Correlación entre rentabilidad predicha y rentabilidad real para DIA.MC	44
Fig. 5.6: Movimiento de las medias de las gaussianas para ACS.MC.....	44
Fig. 5.7: Evolución temporal de las coordenadas u_1 , u_2 , u_3 para cada una de las medias de las gaussianas para ACS.MC	45
Fig. 5.8: Evolución temporal de las tres entradas para cada uno de las 5 gaussianas para ACS.MC.....	46
Fig. 5.9: Evolución de las rentabilidades de cada gaussiana para ACS.MC.....	47
Fig. 5.10: Evolución de las rentabilidades promediadas de cada gaussiana para ACS.MC	47
Fig. 5.11: Estrategia de inversión para ACS.MC.....	48
Fig. 5.12: Estrategia de inversión para ACS.MC.....	48
Fig. 5.13: Estrategia de inversión complementaria para ACS.MC	49
Fig. 5.14: Estrategia de inversión fija para ACS.MC.....	49
Fig. 5.15: Salida por pantalla de la información de la estrategia de inversión para ACS.MC.....	50
Fig. 6.1: Estrategia de inversión para las acciones utilizadas para la optimización multiobjetivo de metaparámetros para K-means	56
Fig. 6.2: Estrategia de inversión para las acciones utilizadas para la validación de metaparámetros en la optimización multiobjetivo para K-means.....	57
Fig. 6.3: Estrategia de inversión para las acciones utilizadas para la optimización restringida de metaparámetros para K-means	59
Fig. 6.4: Estrategia de inversión para las acciones utilizadas para la validación de metaparámetros en la optimización restringida para K-means	60
Fig. 6.5: Estrategia de inversión para las acciones utilizadas para la optimización multiobjetivo de metaparámetros para GMM.....	62
Fig. 6.6: Estrategia de inversión para las acciones utilizadas para la validación de metaparámetros en la optimización multiobjetivo para GMM	63
Fig. 6.7: Estrategia de inversión para las acciones utilizadas para la optimización restringida de metaparámetros para GMM	65
Fig. 6.8: Estrategia de inversión para las acciones utilizadas para la validación de metaparámetros en la optimización restringida para GMM.....	66

ÍNDICE DE TABLAS

Tabla 4.1: Coordenadas de los centroides para ACS.MC en el día 18 de marzo de 2016	27
Tabla 4.2: Información de cada cluster para ACS.MC en el día 18 de marzo de 2016 ...	29
Tabla 5.1: Coordenadas de las medias de las gaussianas para ACS.MC en el día 18 de marzo de 2016.....	41
Tabla 5.2: Información de cada gaussiana para ACS.MC en el día 18 de marzo de 2016	41
Tabla 6.1: Valores de los metaparámetros optimizados	54
Tabla 6.2: Optimización multiobjetivo para K-means	54
Tabla 6.3: Optimización restringida y con racha negativa 10% para K-means	58
Tabla 6.4: Optimización multiobjetivo para GMM	61
Tabla 6.5: Optimización restringida y con racha negativa 10% para GMM	64
Tabla 7.1: Rentabilidades anuales obtenidas en la optimización multiobjetivo para K-means.....	67
Tabla 7.2: Rentabilidades anuales obtenidas en la optimización multiobjetivo para K-means.....	67

1. Introducción

El aprendizaje automático o *machine learning* consiste en diseñar algoritmos que encuentren una tendencia natural o modelo de comportamiento en los datos de un proceso para predecir su evolución futura o tomar mejores decisiones que conduzcan a su optimización. En primer lugar, se seleccionan los datos más relevantes del proceso y se analizan los registros históricos disponibles para esos datos. Una vez identificados los patrones o modelos de comportamiento, se aplica dicho modelo a los nuevos datos con el fin de tomar decisiones acertadas sobre el proceso en cuestión. [1]

En este proyecto, se va a aplicar el aprendizaje automático a las cotizaciones históricas de un conjunto de valores bursátiles con el fin de optimizar su rentabilidad al gestionar una cartera de valores. Se utiliza esta aplicación del aprendizaje automático, entre otros motivos, por la disponibilidad inmediata de un gran volumen de datos, lo cual permite obtener conclusiones sólidas sobre la eficacia de los métodos empleados.

La aplicación del aprendizaje automático a la Bolsa permite, a partir del histórico de cotizaciones de cada acción, realizar predicciones más precisas de la cotización futura o identificar patrones que estén asociados a tendencias alcistas en dicha cotización. Estos métodos ayudan a tomar la decisión de cuándo invertir y conocer la dirección que toman los mercados financieros, maximizando así las ganancias obtenidas. Aplicando el aprendizaje automático, un sistema de inversión automático podría evaluar las distintas acciones y elegir dónde asignar los fondos disponibles.

1.1. Motivación

Una de las motivaciones de este proyecto es detectar inversiones rentables a partir de la evolución del precio de un gran volumen de acciones pertenecientes a diferentes mercados internacionales. Se dispone de datos registrados en Yahoo Finance correspondientes a 10 años.

Se ha empleado el aprendizaje automático no supervisado, en particular la técnica de *clustering*, para identificar patrones en la evolución del precio de las acciones que generen inversiones rentables a medio y largo plazo.

Por otro lado, como se ha explicado anteriormente, hay una gran cantidad de aplicaciones del aprendizaje automático y profundizar en el conocimiento de estas técnicas es otra de las motivaciones de este trabajo. Las herramientas desarrolladas en este proyecto podrían ser utilizadas en otros campos.

Además, existe una línea de investigación acerca del aprendizaje automático aplicado a la Bolsa en la universidad, por lo que este proyecto también va orientado a la ampliación de proyectos realizados anteriormente. [19] [20]

1.2. Objetivos

Los objetivos principales de este proyecto son los siguientes:

1. Diseñar un entorno de simulación y seleccionar un criterio (KPI) para evaluar la eficacia de diferentes métodos de parametrización de precios y de clustering en la identificación de patrones rentables.
2. Emplear el entorno anterior para encontrar la mejor combinación de parametrización y clustering y poder optimizar sus metaparámetros según el índice de funcionamiento (KPI) definido previamente.
3. Desarrollar una estrategia de inversión basada en la combinación óptima y evaluar la rentabilidad obtenida en la gestión de una cartera de valores a partir de datos históricos.

1.3. Metodología

Durante el desarrollo del proyecto se han completado las siguientes tareas:

- Estudio de los métodos a emplear: aprendizaje no supervisado y clustering.
- Estudio del mercado de acciones.
- Modelos de parametrización: definir diferentes formas de parametrizar la evolución del precio durante un intervalo temporal.
- Reducción de dimensionalidad: investigar la conveniencia de reducir la dimensión de los vectores de parámetros definidos mediante análisis de componentes principales.
- *Clustering*: aplicación de las diferentes técnicas de *clustering* (en entrenamiento y validación) para identificar patrones relevantes que se repitan en el tiempo.
- Análisis de la rentabilidad de los *clusters* o patrones mediante la definición de indicador KPI.
- Definición de una estrategia de inversión basada en el análisis anterior y optimización de la misma (encontrar los metaparámetros óptimos).
- Test: Evaluar la rentabilidad final de dicha estrategia mediante nuevos datos.
- Análisis de resultados
- Redacción de la memoria

1.4. Recursos

Para el desarrollo del proyecto se han utilizado:

- MATLAB_R2018B
- Datos de las acciones extraídas de Yahoo Finance correspondientes a 10 años comprendidos entre el 19 de marzo de 2009 y el 19 de marzo de 2019 para unas 1300 acciones de 19 mercados bursátiles de todo el mundo.

1.5. Estructura de la memoria

El capítulo 1 introduce el concepto de aprendizaje automático y explica la motivación y objetivos que se han seguido a la hora de desarrollar el proyecto. Los objetivos se basan en desarrollar una estrategia de inversión óptima y evaluar la rentabilidad obtenida. La metodología seguida y los recursos utilizados también se añaden en este capítulo.

El capítulo 2 describe el estado de la técnica. Aquí se profundiza más sobre los tipos de aprendizaje automático que existen: aprendizaje supervisado y no supervisado con una de sus técnicas más importantes, el clustering. Se desarrollan diferentes técnicas de clustering profundizando en K-means y GMM, las cuales se implementan en una aplicación en Matlab.

El capítulo 3 describe la aplicación en la que se ha implementado el proyecto, la cual tiene como objetivo la definición, análisis y optimización de la estrategia de inversión. La estrategia de inversión se basa en tres señales de entradas y una señal de salida que también se desarrollan en este capítulo.

Para explicar cada uno de los procesos que siguen los dos métodos de clustering se utiliza un ejemplo determinado. La técnica K-means es desarrollada en el capítulo 4 mientras que la técnica de GMM se realiza en el capítulo 5.

Una vez se explica el procedimiento de los métodos, en el capítulo 6, se optimizan los metaparámetros utilizados en cada caso, con el fin de desarrollar la mejor estrategia de inversión.

El capítulo 7 describe las conclusiones, aportaciones y futuros desarrollos del proyecto.

2. Estado del arte

El aprendizaje automático se ha aplicado durante los últimos años en muchos ámbitos diferentes ya que permite procesar un gran volumen de datos en un tiempo relativamente corto. Algunos ejemplos de aplicación son los siguientes:

- Muchas plataformas de comercio electrónico utilizan el aprendizaje automático para ofrecer recomendaciones a los usuarios sobre productos en función de sus búsquedas, preferencias o compras anteriores.
- Los móviles más avanzados realizan automáticamente una foto cuando detectan el patrón correspondiente a una sonrisa, mediante el análisis de una imagen. De la misma manera, mediante el reconocimiento de voz es posible identificar la voz del usuario una vez eliminado el ruido de fondo y ejecutar las acciones solicitadas. [2]
- Los mensajes recibidos como *spam* en los programas de gestión del correo electrónico también utilizan técnicas de *machine learning*. El programa es capaz de reconocer aquellos mensajes que no tienen relación con tu bandeja de entrada y los clasifica como tales.
- En la logística de fabricación, almacenamiento, transporte y suministro de bienes de consumo también es habitual la aplicación del aprendizaje automático. Las empresas predicen las tendencias de compra de los consumidores, así como la previsión de la demanda. De esta manera pueden planificarse de forma óptima todas las etapas del negocio. [3]
- En conducción de coches autónomos es posible predecir los movimientos de los peatones combinando el aprendizaje automático y la percepción humana. [2]
- En el sector sanitario se puede mejorar el diagnóstico de ciertas enfermedades a partir de los síntomas del paciente y el resultado de los análisis médicos. [4]

2.1. El mercado bursátil

La Bolsa apareció en el siglo XIV y ha ido evolucionado y perfeccionándose a lo largo de los siglos. En el mercado bursátil, el valor de una empresa se divide en participaciones financieras, cada una de las cuales tienen el mismo valor unitario. Estas participaciones se denominan acciones. Hoy en día, la compra de acciones es una de las inversiones más comunes. Los accionistas adquieren estos títulos con el fin de obtener dividendos

mientras que las compañías emiten estas acciones por la necesidad de financiación. El precio de las acciones evoluciona continuamente, por esa razón los inversores están interesados en predecir dicha evolución para adquirir aquellas acciones que más beneficios les proporcionen. [6]

Existen dos tipos de análisis para poder predecir el comportamiento futuro de las acciones: el análisis fundamental y el análisis técnico.

2.1.1. Análisis fundamental

Según el análisis fundamental, el comportamiento futuro de las acciones depende del estado de la empresa (balance de situación, cuenta de pérdidas y ganancias, liquidez, rentabilidad, gastos...) y del sector al que pertenece. A partir del análisis de estas características y del sector económico donde opera se estiman los beneficios futuros.

2.1.2. Análisis técnico

El análisis técnico trata de estudiar la evolución futura del precio de la acción a partir de los datos históricos de dicho precio. Por tanto, al contrario que el análisis fundamental, el análisis técnico estudia las tendencias del mercado utilizando gráficos de evolución del precio (charts). [7]

Como se ha comentado anteriormente, este proyecto se centra en la aplicación del aprendizaje automático al análisis técnico con el objetivo de predecir el precio futuro de las acciones. Una de las gestoras de fondos más grandes del mundo, Goldman Sachs AM, ya utiliza el aprendizaje automático para obtener información de las tendencias del mercado. [5]

2.2. Tipos de aprendizaje automático

Existen dos tipos de aprendizaje automático: el aprendizaje supervisado y el aprendizaje no supervisado.

2.2.1. Aprendizaje supervisado

El aprendizaje supervisado (*supervised machine learning*) consiste en desarrollar modelos de predicción que relacionan entradas y salidas. Estos modelos permitirán, por ejemplo, predecir las salidas en el futuro a partir de los datos de las entradas disponibles. Este método utiliza la regresión y la clasificación para el desarrollo de los modelos predictivos.

La regresión es una medida continua que establece una relación entre una serie de variables independientes y la variable dependiente que necesitamos predecir. Con la regresión conseguimos crear la función que mejor encaja con los datos proporcionados. Los parámetros se ajustan a los datos del entrenamiento buscando minimizar el error de estimación de las salidas.

En la clasificación, el algoritmo asigna una etiqueta a la salida conocidas las variables de entrada. Para aprender a elegir la etiqueta, el sistema se ha entrenado previamente con muchos datos de los que se conocen tanto las entradas como las salidas.

Existen tres fases en la aplicación del aprendizaje supervisado: entrenamiento, validación y test.

- En el entrenamiento se trabaja con los datos de entrenamiento, o training data, para poder predecir el modelo cuyo error esté minimizado. Es importante no utilizar un modelo muy complejo que se ajuste muy bien al entrenamiento ya que podemos caer en overfitting o sobreaprendizaje [8]. Este modelo no serviría para generalizar en otras situaciones.
- Una vez ajustado el modelo, se utilizan nuevos datos para verificar si el algoritmo funciona y elegir los metaparámetros que minimicen el error, este paso corresponde a la validación. Los metaparámetros consisten en las características que se estudian en los datos, así como otras propiedades como cuántos grupos diferentes se quieren formar. [9]
- Finalmente, se realiza la verificación final (datos de *test*) para comprobar si el modelo está bien identificado. Los datos utilizados en este último paso no pueden haber sido utilizados anteriormente ya que podrían estar “contaminados”. Se suelen utilizar diferentes criterios a la hora de elegir qué proporción de datos se utilizan para cada parte, siendo muy común el uso del 50% de los datos para entrenamiento, 30% para validación y 20% para el test.

2.2.2. Aprendizaje no supervisado

En el aprendizaje no supervisado (*unsupervised machine learning*) el sistema no dispone previamente de criterios con los que clasificar los datos. En este caso solo se conocen los datos de entrada y necesitamos agrupar éstos para poder predecir su comportamiento futuro. Es muy importante normalizar todos los datos para poder relacionarlos.

Una de las técnicas más importantes de aprendizaje no supervisado es el **clustering**, la cual trata de organizar en grupos los datos de entrada que se parecen de alguna manera. Estos grupos son denominados clusters. Un inconveniente del clustering es que no existe un criterio definido para reconocer si un cluster es bueno o no. Es el usuario quien tiene que decidir sobre la bondad de cada cluster [10]. El clustering tiene infinidad de aplicaciones como el análisis de redes sociales, la segmentación del mercado (marketing), la agrupación de búsquedas, etc.

Existen diferentes tipos de técnicas de clustering como K-means, Gaussian Mixture Models (GMM) o redes neuronales, entre otros.

- K-means: es un algoritmo iterativo que trata de organizar y clasificar los datos (puntos en un espacio de n dimensiones) en torno a atractores o centroides en cada iteración. Una vez elegido el número de clusters o centroides que se quieren identificar, se sitúan aleatoriamente los centroides iniciales de estos *clusters*. El objetivo de esta técnica es iterar hasta conseguir minimizar la suma de las distancias de los datos a sus centroides más próximos [10]. A la hora de definir la distancia entre los datos y los centroides se pueden utilizar diferentes métodos, siendo el más habitual la distancia euclídea.

La técnica K-means se ha aplicado, por ejemplo, sobre los rendimientos y valoración de 106 acciones en diferentes momentos de la Bolsa de Bombay. [11] Para ello se utilizaron datos de empresas del año fiscal 2007-2008 de Capitaline Databases Plus (una base de datos de información financiera). El clustering se aplicó sobre las devoluciones (1 día, 1 semana, 30 días...) y ratios de validación (Precio al valor contable, ganancia del precio, etc.). Se realizó la extracción de información significativa para formar una cartera de acciones y así minimizar el riesgo de inversión gracias a esta diversificación.

Otra de sus aplicaciones consiste en identificar los factores que impulsan las actuaciones de los inversores más activos. [12] En esta referencia, se estudiaron datos de 120 acciones incluidos los datos sobre si el comprador/vendedor era de alta frecuencia. Se utilizaron datos de 3 meses de NASDAQ HFT.

- GMM: este método utiliza un modelo de suma de distribuciones gaussianas n -dimensionales sobre los datos y ajusta los parámetros de las gaussianas (medias y matriz de covarianzas) para que se adapten ellos. Las gaussianas se ajustan automáticamente en las zonas donde hay más acumulación de datos de tal manera que sus medias se concentran donde existe mayor densidad de puntos.

Después, a cada punto se le asocia un cluster, que será la gaussiana a la que tenga mayor grado de pertenencia. Por lo tanto, existe el mismo número de gaussianas que de clusters definidos. [13]

Debido a la complejidad de este modelo, éste se ralentiza cuando la dimensionalidad es alta. Por ello, es conveniente utilizar una reducción de dimensionalidad previa.

- Redes neuronales: Existe un tipo de redes neuronales denominado Self-Organized Map (SOM) que se aplica en aprendizaje no supervisado. Este método está formado por dos capas de neuronas, la capa de entrada, la cual recibe los datos, y la capa de salida, la cual organiza la información, reduce la dimensionalidad y extrae las características principales. [14]

La técnica de redes neuronales se aplica, por ejemplo, para pronosticar el valor de los precios futuros en el mercado de valores. [15] En esta referencia, se demuestra que estos métodos superan los métodos tradicionales de pronóstico de precios. Se dispone de datos de S&P 500 durante 361 meses, desde el 1 de enero de 1980 hasta el 1 de enero de 2009. Éstos incluyen precios anteriores, volumen, precios del productor, niveles de producción industrial, indicadores técnicos, etc.

Este tipo de aprendizaje automático también dispone de las 3 fases: entrenamiento, validación y test.

En la literatura técnica se aplican otros métodos de clustering a la inversión en Bolsa, por ejemplo, price clustering para predecir la tendencia de los precios de las acciones a agruparse en números redondos (10,20,30, etc.). [16] Para ello se dispone de datos del mercado de valores holandés entre los años 1990-2001.

Otra aplicación de clustering consiste en predecir el porcentaje de cambio en el precio futuro de las acciones de una compañía. [17] Se dispone de datos sobre los datos del precio de cierre de las acciones de 85 empresas del S&P 500 durante 126 días. Los datos corresponden al periodo existente entre enero de 1996 y diciembre de 2015.

También se aplica el clustering sobre datos del mercado de valores de Toronto para el diseño y creación del flujo de trabajo paralelo en las empresas financieras con la finalidad agrupar datos de una cartera de acciones con un tiempo de respuesta mínimo. [18]

En la universidad se han desarrollado otros proyectos de aprendizaje automático aplicado a la Bolsa, utilizando tanto aprendizaje supervisado [19] como aprendizaje no supervisado aplicando K-means y GMM. [20]

3. Descripción de la aplicación

La aplicación, mostrada en la figura 3.1, tiene como objetivo la definición, análisis y optimización de estrategias de inversión aplicadas sobre un valor o acción concreta. Para este fin se dispone de datos procedentes de Yahoo Finance correspondientes a 10 años para unas 1300 acciones de diferentes mercados bursátiles de todo el mundo.

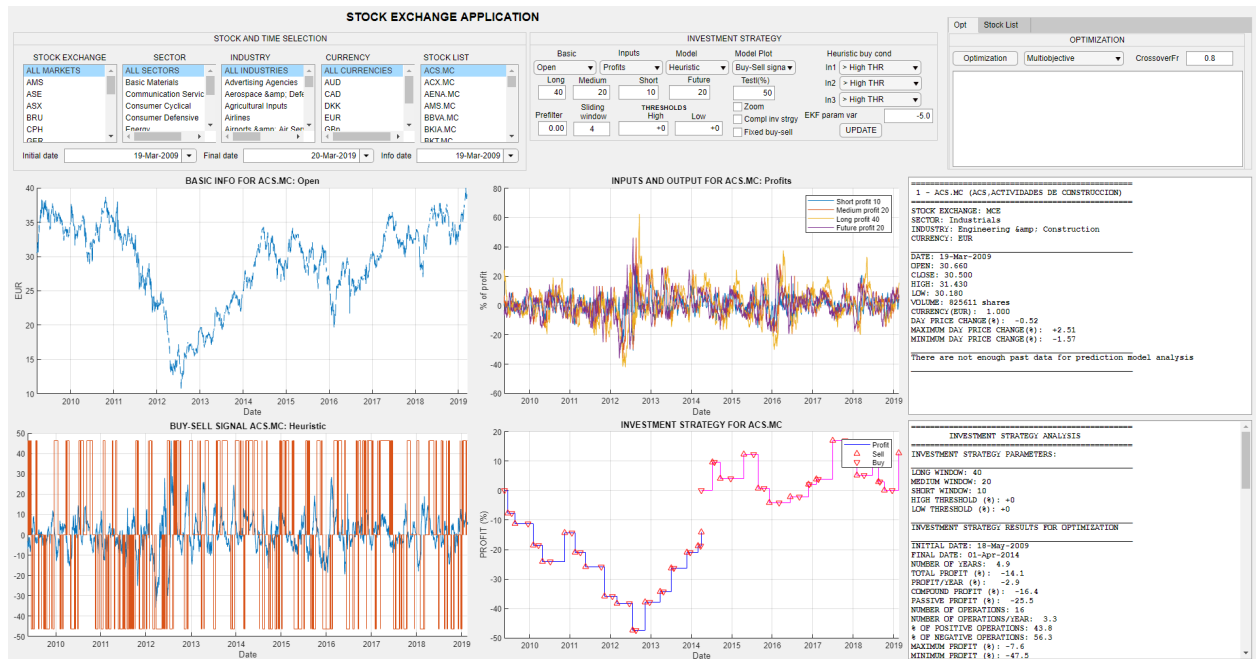


Fig. 3.1: Interfaz de la aplicación utilizada

Como muestra la figura 3.2, en la aplicación se puede seleccionar la acción, teniendo en cuenta su mercado, sector, industria o moneda, cuya estrategia de inversión se quiere optimizar. Además, también se puede seleccionar la fecha con las que se quiere trabajar en los 10 años disponibles.

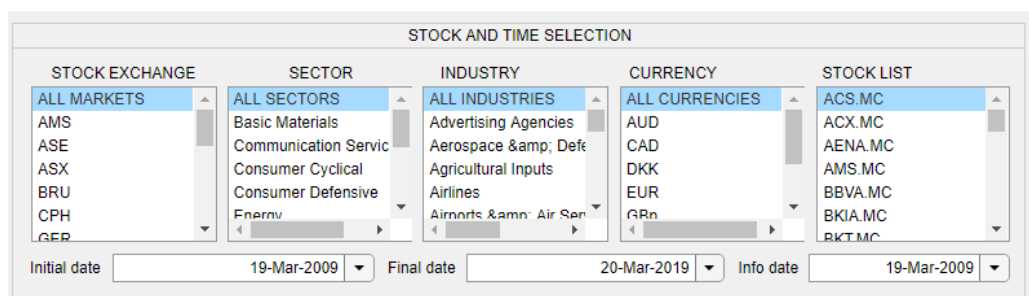


Fig. 3.2: Función de selección de las acciones

La figura 3.3 muestra la elección de las entradas con las cuales se quiere trabajar y que serán explicadas más adelante en este capítulo. También aparece la selección del tipo de modelo que se quiere utilizar para definir la estrategia de inversión.

Fig. 3.3: Configuración de opciones de la estrategia de inversión

Además, se podrán optimizar los metaparámetros utilizados en el proceso como se puede observar en la figura 3.4 y que será aplicado en el capítulo de análisis de resultados sobre una lista de valores previamente seleccionados.

Fig. 3.4: Configuración de optimización

Cada una de las gráficas de la figura 3.5 forma parte del proceso que será explicado en éste y posteriores capítulos dependiendo del método utilizado. A la derecha de la figura aparece información relevante de cada una de las gráficas que ayudarán a analizar la estrategia de inversión en cada caso.



Fig. 3.5: Gráficas e información de la estrategia de inversión

La estrategia de inversión de basa en tres señales de entrada generadas de distintas formas a partir de las cotizaciones más recientes correspondientes a cada sesión. Cada sesión la identificamos con un índice k que define el orden temporal en el intervalo de los 10 años. Si $k=100$, se trata de la sesión número 100 dentro de ese intervalo.

Las señales de compra y venta de una estrategia de inversión pueden generarse de dos formas distintas:

- Mediante un procedimiento heurístico que utiliza condiciones lógicas basadas en umbrales y cruces entre las señales de entrada.
- Mediante un modelo de predicción de la rentabilidad futura. Algunos de esos modelos de predicción se basan en clustering o identificación de patrones.

Se trabaja con precios o cotizaciones normalizadas. Para eso se define en cada sesión k una ventana con las cotizaciones más recientes del valor que se esté analizando. Los precios normalizados para esa ventana se calculan como (k es el dato de la sesión actual y N la longitud de esa ventana):

Precio de cierre normalizado:

$$C_N[k-i] = \frac{(\text{close}[k-i] \text{ currency}[k-i])}{(\text{open}[k-N+1] \text{ currency}[k-N+1])} - 1$$

siendo $i=0\dots, N-1$

Precio de apertura normalizado:

$$O_N[k-i] = \frac{(\text{open}[k-i] \text{ currency}[k-i])}{(\text{open}[k-N+1] \text{ currency}[k-N+1])} - 1$$

siendo $i=0\dots, N-1$

siendo $\text{close}[k]$ el precio de cierre en la sesión k , $\text{open}[k]$ el precio de apertura en la sesión k y $\text{currency}[k]$ el cambio de moneda en la que se cotiza la acción con respecto al euro.

De esta forma, el precio normalizado para cada sesión es la cotización en unitarias de la acción en esa ventana deslizante de N datos, teniendo también en cuenta la tasa de cambio de la moneda en la que se cotiza la acción con respecto al euro. Como base para el cálculo de precio en unitarias, se toma el precio en euros al comienzo de la ventana. Por lo tanto, el precio normalizado depende de la longitud N de la ventana que se escoja.

Además, ese precio normalizado se puede filtrar (es una opción disponible en la aplicación) mediante un filtro paso bajo simétrico (se pasa en ambos sentidos de la

ventana para evitar retrasos) con fin de suavizar el precio normalizado y reducir el nivel de ruido presente en el mismo.

Se va a representar un ejemplo utilizando la acción ACS.MC en la sesión 1000. La longitud de la ventana larga (L) es 40 días, la ventana mediana (M) 20 días y la ventana corta (S) 10 días.

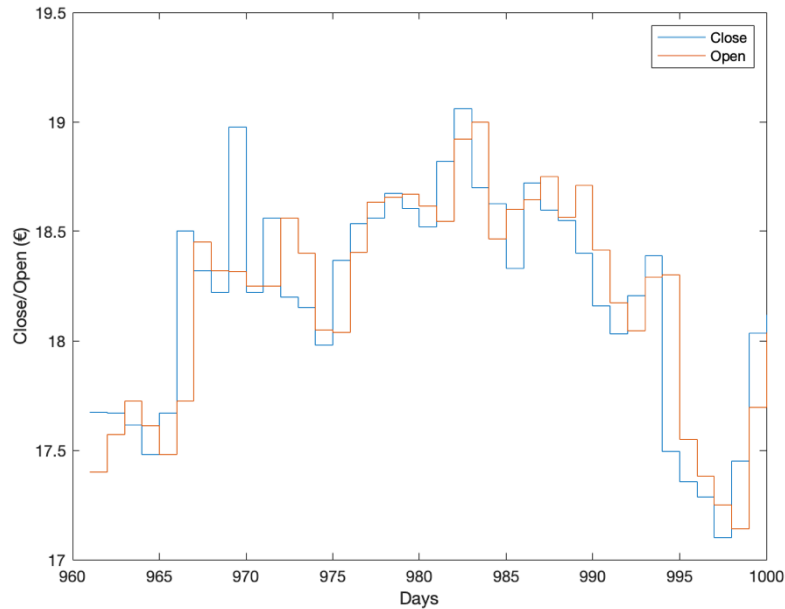


Fig. 3.6: Precio de cierre y apertura para ACS.MC

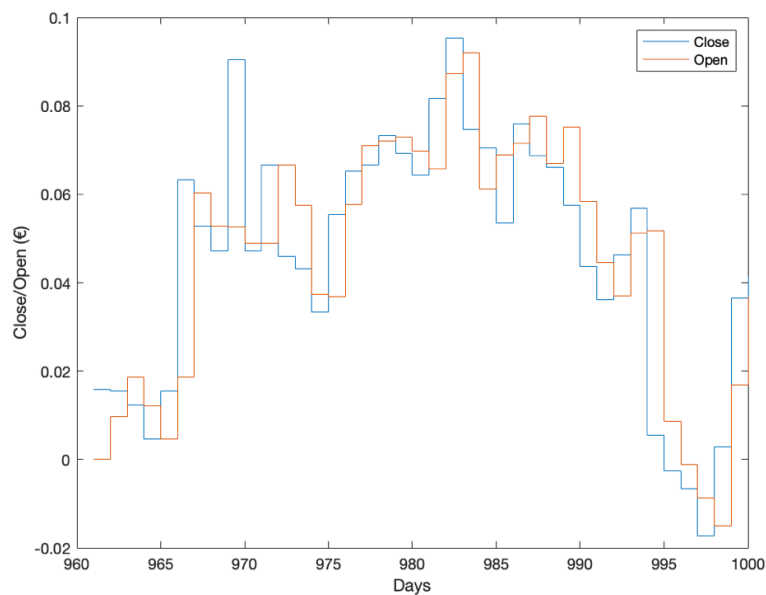


Fig. 3.7: Precio de cierre y apertura normalizados para ACS.MC

3.1. Entradas y salidas al modelo de predicción

Existen distintos tipos de entradas para generar la estrategia de inversión.

- Rentabilidades: Las tres entradas se calculan a partir de los precios normalizados o rentabilidades en tres ventanas distintas, una de longitud larga (L), otra de longitud media (M) y otra de longitud corta (S). De esta forma, las tres entradas serían:

$$u1[k] = 100 \frac{(C_N[k] - O_N[k - L + 1])}{1 + O_N[k - L + 1]}$$

$$\text{Si } N=L, \text{ entonces } u1[k] = 100C_N[k]$$

$$u2[k] = 100 \frac{(C_N[k] - O_N[k - M + 1])}{1 + O_N[k - M + 1]}$$

$$u3[k] = 100 \frac{(C_N[k] - O_N[k - S + 1])}{1 + O_N[k - S + 1]}$$

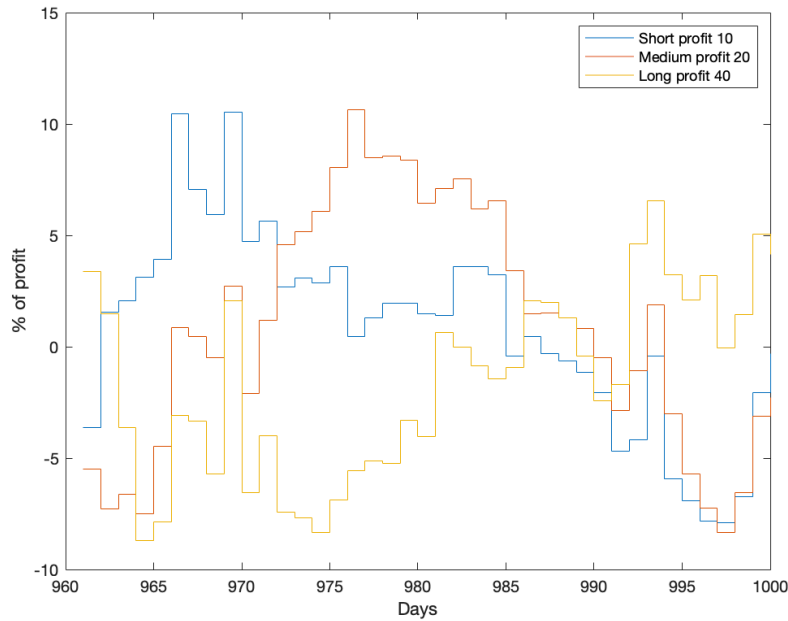


Fig. 3.8: Rentabilidades para ACS.MC

- Diferencias de rentabilidades: las tres entradas se calculan a partir de las diferencias de las rentabilidades en tres ventanas distintas, una de longitud larga (L), otra de longitud media (M) y otra de longitud corta (S) calculadas en el caso anterior. De esta forma, las tres entradas serían:

$$u1[k] = 100 \frac{(C_N[k] - O_N[k - L + 1])}{1 + O_N[k - L + 1]}$$

$$u2[k] = 100 \left(\frac{(C_N[k] - O_N[k - L + 1])}{1 + O_N[k - L + 1]} - \frac{(C_N[k] - O_N[k - M + 1])}{1 + O_N[k - M + 1]} \right)$$

$$u3[k] = 100 \left(\frac{(C_N[k] - O_N[k - M + 1])}{1 + O_N[k - M + 1]} - \frac{(C_N[k] - O_N[k - S + 1])}{1 + O_N[k - S + 1]} \right)$$

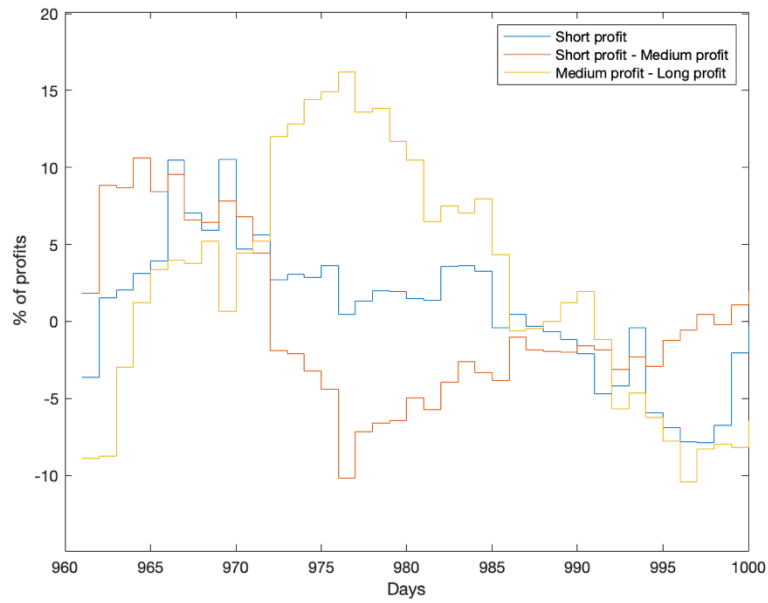


Fig. 3.9: Diferencias de rentabilidades para ACS.MC

- Rentabilidad más reciente, la rentabilidad máxima y la rentabilidad mínima para una longitud de ventana determinada. Las tres entradas serían:

$$u1[k] = 100 C_N[k]$$

$$u2[k] = 100 \max(C_N[i]), \text{ siendo } i = k-N+1 \dots, k$$

$$u3[k] = 100 \min(C_N[i]), \text{ siendo } i = k-N+1 \dots, k$$

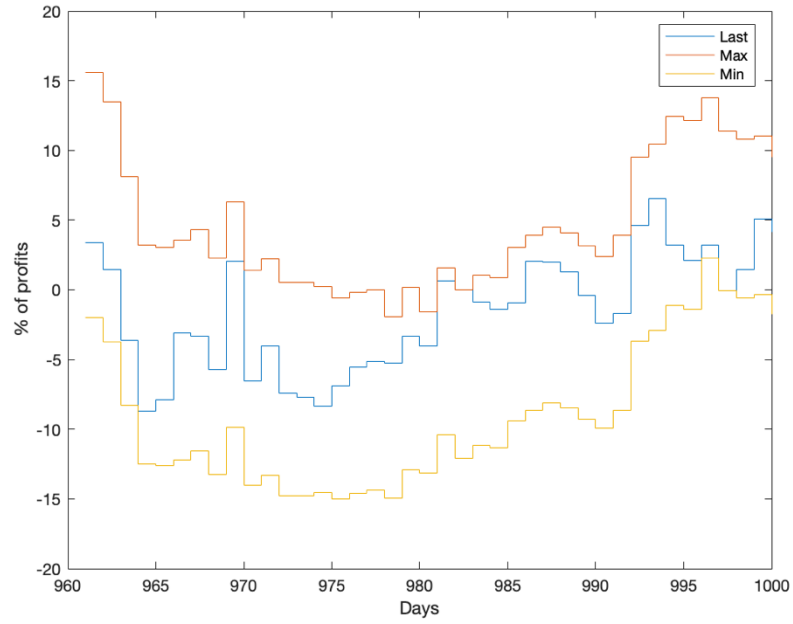


Fig. 3.10: Rentabilidad más reciente, máxima y mínima para ACS.MC

- Diferencias entre la rentabilidad más reciente, la rentabilidad máxima y la rentabilidad mínima calculadas anteriormente. En este caso, las tres entradas serían las siguientes:

$$\begin{aligned}
 u1[k] &= 100 C_N[k] \\
 u2[k] &= 100 (\max(C_N[i]) - \min(C_N[i])), \text{ siendo } i = k-N+1 \dots, k \\
 u3[k] &= 100 (C_N[k] - \min(C_N[i])), \text{ siendo } i = k-N+1 \dots, k
 \end{aligned}$$

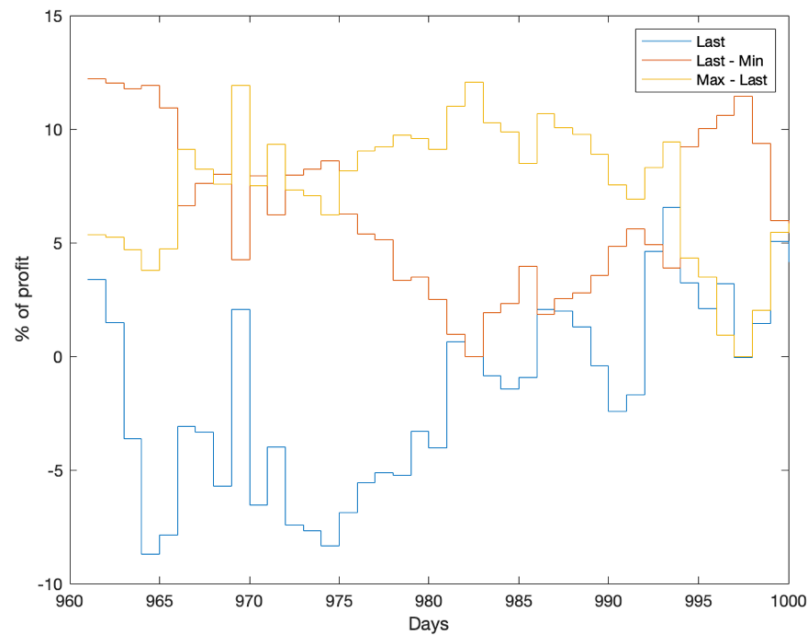


Fig. 3.11: Diferencias rentabilidad más reciente, máxima y mínima para ACS.MC

- Medias de las rentabilidades en tres ventanas distintas, una de longitud larga (L), otra de longitud media (M) y otra de longitud corta (S). Las entradas serían:

$$u1[k] = 100 \frac{\sum_{i=k-L+1}^{i=k} C_N[k]}{L}$$

siendo $i = k-L+1, \dots, k$

$$u2[k] = 100 \frac{\sum_{j=k-M+1}^M C_N[k]}{M}$$

siendo $j = k-M+1, \dots, k$

$$u3[k] = 100 \frac{\sum_{h=k-S+1}^S C_N[k]}{S}$$

siendo $h = k-S+1, \dots, k$

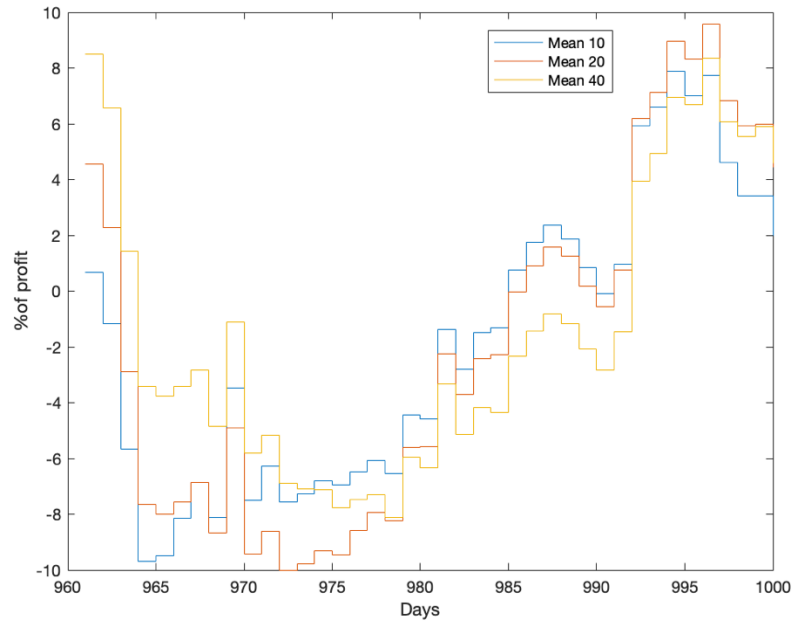


Fig. 3.12: Medias de las rentabilidades para ACS.MC

- Diferencias entre las medias de las rentabilidades en tres ventanas distintas, larga (L), media (M) y corta (S) y la rentabilidad más reciente de la ventana.

$$u1[k] = 100 \left(P_N[k] - \frac{\sum_{i=k-L+1}^{i=k} C_N[k]}{L} \right)$$

$$u2[k] = 100 \left(\frac{\sum_{i=k-L+1}^{i=k} C_N[k]}{L} - \frac{\sum_{k-M+1}^M C_N[k]}{M} \right)$$

$$u3[k] = 100 \left(\frac{\sum_{k-M+1}^M C_N[k]}{M} - \frac{\sum_{k-S+1}^S C_N[k]}{S} \right)$$

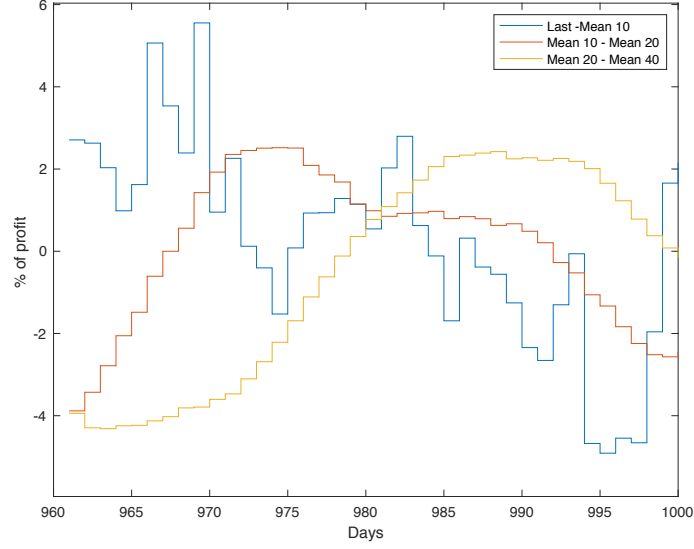


Fig. 3.13: Diferencias de las medias de rentabilidades para ACS.MC

- Regresión lineal: se realiza una regresión lineal con las rentabilidades de la ventana larga para ajustar con una función cuadrática:
 $y = a_2 i^2 + a_1 i + a_0$, siendo $i = 1, \dots, N$

Con ello se genera otro tipo de entradas que son las rentabilidades calculadas por separado de cada uno de los 3 términos de la función cuadrática.

Por tanto, las entradas son:

$$\begin{aligned} u1[k] &= 100 a_0 \\ u2[k] &= 100 a_1 N \\ u3[k] &= 100 a_2 N^2 \end{aligned}$$

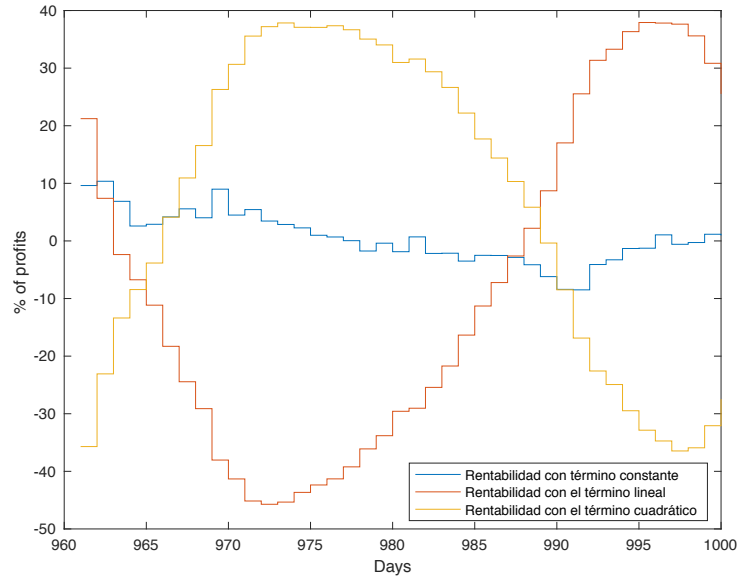


Fig. 3.14: Regresión lineal para ACS.MC

- Rentabilidades muestreadas: Comprando al principio de la ventana larga, cada una de las entradas refleja la rentabilidad a determinados días: la rentabilidad pasados los días que forma a ventana corta (S), la ventana media (M) y la ventana larga (L).

Normalmente los valores de N y L coinciden.

$$u1[k] = 100 \ C_N[k - N + L]$$

$$\text{Si } N=L, \text{ entonces } u1[k] = 100 \ C_N[k]$$

$$u2[k] = 100 \ C_N[k - N + M]$$

$$u3[k] = 100 \ C_N[k - N + S]$$

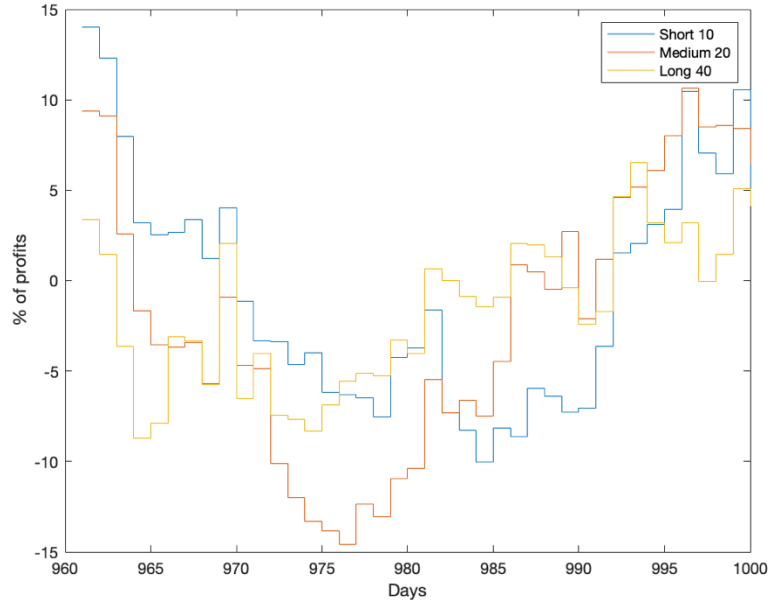


Fig. 3.15: Rentabilidades muestreadas para ACS.MC

- Diferencias entre las rentabilidades muestreadas:

$$u1[k] = 100 C_N[k - N + L]$$

$$u2[k] = 100 (C_N[k - N + L] - C_N[k - N + M])$$

$$u3[k] = 100 (C_N[k - N + M] - C_N[k - N + S])$$

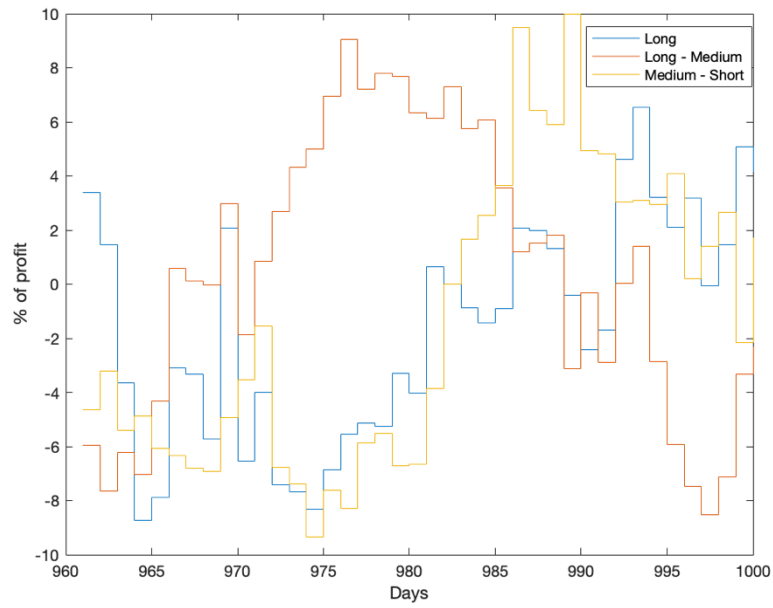


Fig. 3.16: Diferencia entre rentabilidades muestreadas para ACS.MC

Por otro lado, la salida de la función es la rentabilidad futura y . Esta rentabilidad se calcula de acuerdo a la longitud de una ventana futura.

3.2. Modelo de predicción

Para realizar la estrategia de inversión el objetivo es crear una función matemática F que relacione las tres entradas con la salida. Esta función depende, por tanto, de las entradas u_1 , u_2 y u_3 y predice la rentabilidad futura y :

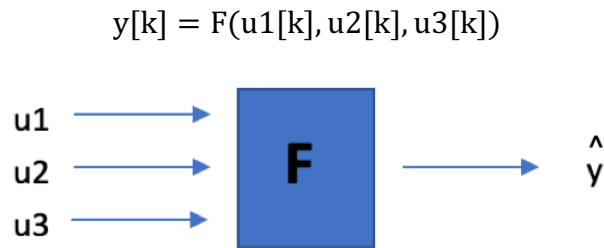


Fig. 3.17: Función matemática del modelo de predicción

En este caso, el modelo de predicción se realizará de acuerdo a la técnica de clustering.

En primer lugar, se define una ventana de N datos previa al día actual. Cada punto de esa ventana constituye un punto en el espacio 3D debido a que dispone de tres coordenadas: u_1 , u_2 y u_3 . De cada uno de estos puntos también conocemos su rentabilidad futura y .

En segundo lugar, se agrupan estos puntos en subconjuntos utilizando diferentes criterios los cuales se explicarán más en detalle en posteriores capítulos.

- El método K-means agrupa los datos entorno a unos centroides seleccionados con el fin de minimizar la suma de las distancias euclideas a los puntos más cercanos del conjunto de datos.
- El método GMM utiliza gaussianas con mayor densidad en las zonas donde hay más acumulación de puntos. Cada punto pertenece a cada gaussiana con una determinada probabilidad.

Una vez realizadas las agrupaciones, se calcula la rentabilidad futura media que se obtuvo con todos los puntos que pertenecen a cada grupo y ésta es la rentabilidad que se le asigna a cada una de las agrupaciones.

Finalmente, para realizar la predicción, se toma el día actual con sus respectivas entradas (u_1 , u_2 y u_3) y según éstas se ve a qué grupo pertenece y se le asigna la rentabilidad de ese grupo como predicción de rentabilidad futura.

Una vez se asigna la predicción de rentabilidad futura al día actual, el modelo se mueve un día y vuelve a tomar la misma longitud de datos previos. Con lo cual, sale un día y entra otro día en la ventana deslizante.

Con los nuevos puntos se modifican las agrupaciones anteriores consiguiendo unas nuevas acorde a los nuevos datos. Estas agrupaciones no se verán muy modificadas debido a que se trata de un cambio de un día frente a una ventana de longitud mucho mayor.

De la misma manera que anteriormente, el día actual pertenecerá a una de esas agrupaciones y le asignaremos su rentabilidad como predicción futura.

Una vez asignada la predicción futura al día actual, el modelo se vuelve a mover un día y así sucesivamente. Por lo tanto, este es un modelo deslizante que utiliza la misma parte de datos para entrenamiento y validación.

3.3. Estrategia de inversión

Una vez obtenidas las predicciones de rentabilidad futura de todos los días, se genera una señal de compra-venta que, como su nombre indica, señala cuando se debe comprar y vender la acción a partir de un test de umbral. Éste establece los umbrales de rentabilidad predicha a partir de los cuales la acción debe comprarse o venderse. Se compra si la rentabilidad está por encima del umbral superior y se vende si la rentabilidad predicha está por debajo del umbral inferior.

Existen diferentes estrategias de inversión:

- Utilizar la información de la señal de compra-venta para comprar y vender.
- Utilizar la estrategia complementaria a la señal de compra-venta: comprar cuando la señal indica vender y vender cuando la señal indica comprar.
- Comprar si la rentabilidad predicha está por encima del umbral superior y vender siempre después de la longitud de la ventana futura utilizada para hacer la predicción.

En la estrategia de inversión se distinguen dos intervalos como muestra la figura 3.18. El primer intervalo temporal es para optimizar los metaparámetros. En el segundo

intervalo temporal se mantienen los metaparámetros constantes y se evalúa la estrategia de inversión.

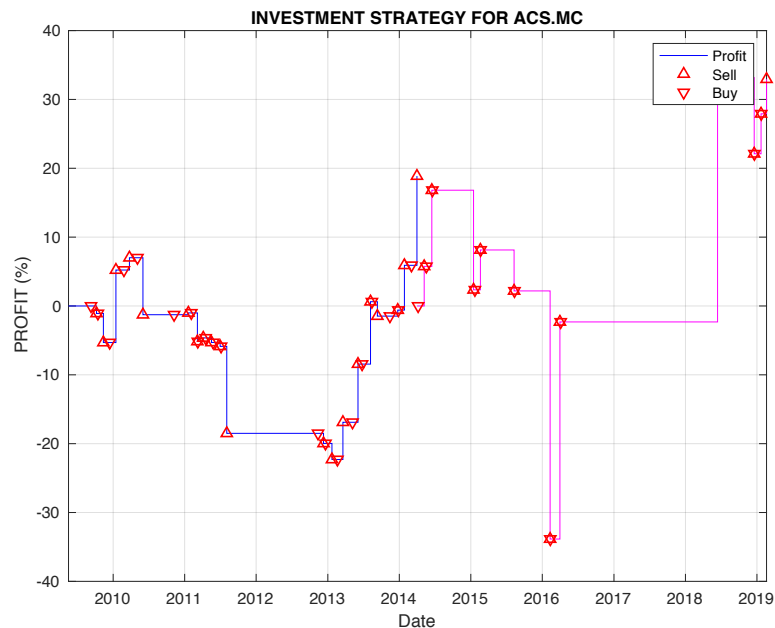


Fig. 3.18: Ejemplo de estrategia de inversión

3.4. Ejemplo utilizado para la ilustración del proceso para los métodos K-means y GMM

Se va a utilizar un ejemplo para ilustrar el procedimiento para K-means y GMM con el fin de poder comparar ambos métodos. A continuación, se detallan los datos comunes de ambos procesos, en los cuales los metaparámetros no han sido optimizados.

Se dispone de datos correspondientes a 10 años, desde el 19 de marzo de 2009 hasta el 19 de marzo de 2019.

Se va a seguir utilizando la acción ACS.MC representada anteriormente asignando como día actual el 18 de marzo de 2016. En la figura 3.19 se definen los datos de esta acción, precio de apertura, precio de cierre, precio máximo y mínimo de la sesión, cambio de la moneda en la que se cotiza la acción con respecto al euro, cambio de precio en el día y su máximo y mínimo. También se indica el número de transacciones de compra-venta realizados durante esa sesión.

```

1 - ACS.MC (ACS,ACTIVIDADES DE CONSTRUCCION)
=====
STOCK EXCHANGE: MCE
SECTOR: Industrials
INDUSTRY: Engineering & Construction
CURRENCY: EUR

DATE: 18-Mar-2016
OPEN: 26.005
CLOSE: 26.550
HIGH: 26.640
LOW: 25.835
VOLUME: 1141393 shares
CURRENCY(EUR): 1.000
DAY PRICE CHANGE(%): +2.10
MAXIMUM DAY PRICE CHANGE(%): +2.44
MINIMUM DAY PRICE CHANGE(%): -0.65

```

Fig. 3.19: Salida por pantalla del precio de cierre y apertura para ACS.MC en el día 18 de marzo de 2016

La longitud de la ventana larga es 40 días, la ventana media consta de 20 días y la ventana corta de 10 días. La predicción de la rentabilidad será a 20 días desde el día actual.

En primer lugar, se selecciona el tipo de entradas al modelo de predicción. En este caso se van a utilizar las rentabilidades correspondientes a 40, 20 y 10 días (u_1 , u_2 , u_3). En la figura 3.20 están representadas las entradas de cada día y rentabilidad futura a 20 días.

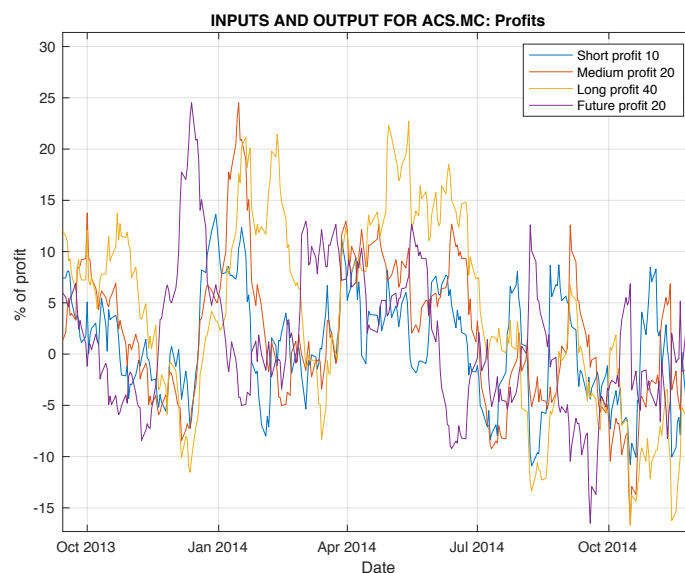


Fig. 3.20: Entradas y salidas del modelo de predicción para ACS.MC

Se utiliza una ventana deslizante de 4 años anteriores al día actual. Cada día de estos 4 años actual, es decir, aproximadamente del 25 de febrero del 2012 al 25 de febrero del

2016 constituye un punto de tres coordenadas, siendo estas cada una de las entradas (u_1 , u_2 , u_3), representados en la figura 3.21. No se han tomado los inmediatamente anteriores 20 días al 18 de marzo del 2016 debido a que de esos días no se tiene la rentabilidad futura a 20 días.

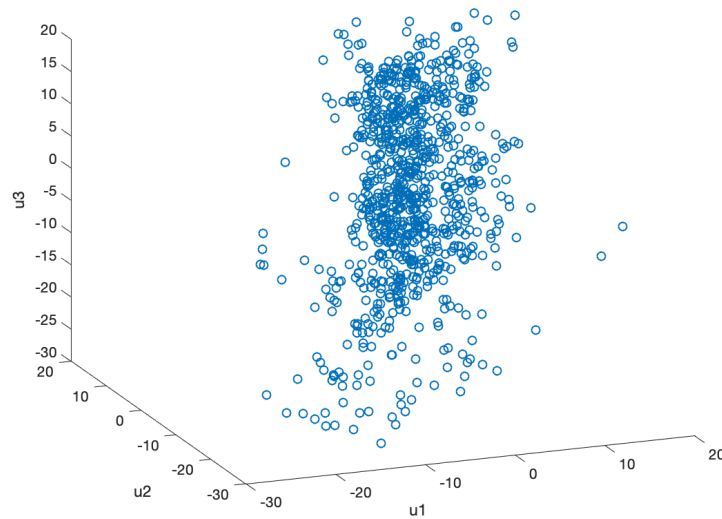


Fig. 3.21: Entradas de cada uno de los días que constituyen los 4 años para ACS.MC

Estos son los datos comunes de los que se dispone correspondientes a ambos métodos. A partir de ahora se va a seguir desarrollando la estrategia de inversión para K-means y GMM en los capítulos 4 y 5 respectivamente.

4. K-means

Como se ha explicado en capítulos anteriores, el método K-means organiza y clasifica los datos entorno a unos centroides. Esta agrupación se basa en la minimización de la distancia euclídea de los datos a sus centroides más próximos.

Con los puntos ilustrados en la figura 3.21 se forman 5 grupos denominados clusters de tal manera que se minimice la suma de las distancias de cada uno de los datos a sus centroides más próximos. En la tabla 4.1 se definen las coordenadas de cada uno de los 5 clusters formados con estos 4 años anteriores al día 18 de marzo de 2016 y en la figura 4.1 quedan representados estos centroides.

CENTROIDES			
	Componente x	Componente y	Componente z
Centroide 1	-9.26	-16.16	-26.73
Centroide 2	+5.48	+11.71	+18.28
Centroide 3	+1.61	+2.97	+6.04
Centroide 4	+17.48	+28.33	-1.33
Centroide 5	-2.48	-5.07	-6.39

Tabla 4.1: Coordenadas de los centroides para ACS.MC en el día 18 de marzo de 2016

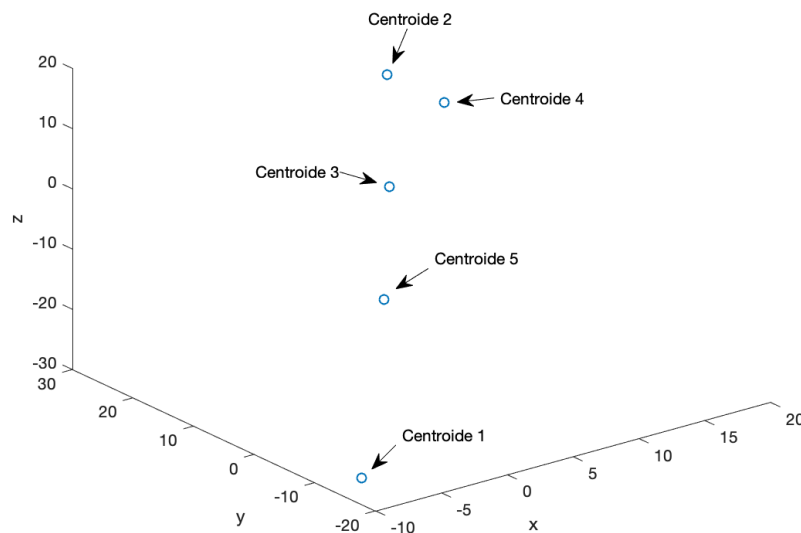


Fig. 4.1: Coordenadas de los centroides para ACS.MC en el día 18 de marzo de 2016

La rentabilidad futura que se asigna a cada cluster es la media de las rentabilidades de cada uno de los puntos que pertenecen a cada agrupación.

En la tabla 4.2 aparece la información de cada uno de los 5 clusters formados:

- Población: porcentaje de puntos que pertenecen a cada cluster.

$$\text{pop}_j = 100 \frac{\sum_{p_i \in c_j} p_i}{N_{p_j}}$$

Donde c_j corresponde a cada uno de los clusters, con $j=1\dots,5$ y P_n son todos los puntos pertenecientes al entrenamiento con p_i cada punto específico, siendo $i=1\dots, P_n$.

- Dispersión de los puntos dentro de cada cluster: indica la concentración de los clusters.

$$i_{\text{media}} = \frac{\sum_{p_i \in c_j} i}{N_{p_j}}$$

$$\text{disp}_j = \sqrt{\frac{\sum_{p_i \in c_j} (i - i_{\text{media}})^2}{N_{p_j}}}$$

- Rentabilidad del cluster: calculado a partir de las medias de las rentabilidades de los puntos que pertenecen a cada cluster.

$$\text{profit}_j = \frac{\sum_{p_i \in c_j} \text{profit}(p_i)}{N_{p_j}}$$

- Rentabilidad promediada del cluster: calculada igual que la rentabilidad del cluster, pero dando más peso a las rentabilidades de los puntos más cercanos a cada centroide.

$$\text{dist}_{p_i \in c_j} = \sum (c_j - p_i)^2$$

$$\text{part}_{i \in c_j} = \frac{1}{\text{dist}_{p_i \in c_j}}$$

$$\text{wprofit}_j = \frac{\sum_{p_i \in c_j} \text{profit}(p_i) * \text{part}_j(p_i)}{\sum_{p_i \in c_j} \text{part}_j(p_i)}$$

- Volatilidad 1 o volatilidad interna: En primer lugar, se forma una recta del 0 al punto de rentabilidad final del cluster (media de todos los puntos). A continuación, se toma la ventana de 20 días de la rentabilidad para ver cómo ha ido evolucionando ésta y se calcula la varianza de cada una de las ventanas con respecto a la recta. La volatilidad interna es la media de estas varianzas.

$$vol_1 = \frac{\sum \text{var}(\text{profit}(f_i))}{N_f}$$

Donde N_f son todos los puntos que pertenecen al intervalo de rentabilidad futura, en este caso 20 días, con f_i cada punto específico, siendo $i=1\dots, N_f$.

- Volatilidad 2 o volatilidad externa: La rentabilidad final del cluster dispone de una media y una desviación típica. Por lo tanto, la desviación típica del conjunto de puntos asociados a un cluster es la volatilidad externa.

$$vol_2 = \sqrt{\frac{\sum (\text{profit}(f_i) - \text{profit}_j)^2}{N_f}}$$

- KPI: es una medida de calidad de cada cluster. Cuando mayor sea el KPI mejor calidad tiene el cluster.

$$KPI = \frac{(\text{wprofit}_j - \text{operation_costs})}{\text{future_window_profit} \text{ pop}_j \text{ disp}_j (1 + vol_1)(1 + vol_2)}$$

$$\text{con } (\text{wprofit} - \text{operation_costs}) > 0$$

Donde wprofit_j es la rentabilidad promediada de cada cluster, operation_costs son los gastos de cada operación, $\text{future_window_profit}$ corresponde a la ventana de predicción de rentabilidad futura, pop_j es la población de cada cluster, disp_j es la dispersión de cada cluster y vol_1 y vol_2 son las volatilidades interna y externa respectivamente.

CLUSTER INFO						
PPL (%)	DSP (%)	PRF (%)	WPR (%)	IVL (%)	FVL (%)	KPI
9.0	35.7	+2.4	+0.6	5.3	18.1	0.00
19.8	20.0	-2.9	-3.0	3.0	8.1	0.00
34.3	18.5	+0.5	+0.3	3.0	7.2	0.00
1.3	3.8	+8.4	+8.9	4.4	4.7	0.06
35.6	22.0	+3.1	+2.8	3.7	10.7	1.28

Tabla 4.2: Información de cada cluster para ACS.MC en el día 18 de marzo de 2016

El día actual, 18 de marzo de 2016, tiene los siguientes datos de rentabilidad en las tres ventanas:

```

INPUTS FOR PREDICTION MODEL:
SHORT PROFIT FOR THE LAST 10 DAYS(%):    3.87
MEDIUM PROFIT FOR THE LAST 20 DAYS(%):   24.36
LONG PROFIT FOR THE LAST 40 DAYS(%):    16.63

```

Fig. 4.2: Salida por pantalla de los datos de las señales de entrada para ACS.MC

En la figura 4.3 aparece la información del día actual, sus 3 entradas, la distancia del punto a cada centroide, la probabilidad con la que pertenece a cada cluster, la rentabilidad futura asignada y la rentabilidad futura promediada del cluster al que pertenece.

Como se puede observar en la figura 4.3, el dato actual de 3 coordenadas pertenece al cluster 2. Por tanto, como predicción de rentabilidad futura se le asigna la rentabilidad de ese cluster, siendo ésta -2.9%.

En este caso, current weighted profit representa la rentabilidad promediada del cluster al que pertenece haciendo la media de la participación del punto en cada uno de los clusters:

$$\text{current_weighted_profit} = \frac{\sum_j \text{profit}_j(p_k) * \text{part}_j(p_k)}{\sum_j \text{part}_j(p_k)}$$

Donde p_k corresponde al día actual.

```

CURRENT PARAMETER VECTOR INFO
PARAMETER VECTOR(%): [ +3.87 +24.36 +16.63]
DISTANCES(%): [60.8 12.9 24.0 22.9 37.9]
MEMBERSHIP(%): [ 8.0 37.8 20.2 21.2 12.8]
CURRENT CLUSTER: 2
CURRENT MEAN PROFIT(%): -2.90
CURRENT WEIGHTED PROFIT(%): +1.38
CURRENT SCALED MEAN PROFIT(%): -22.27
CURRENT SCALED WEIGHTED PROFIT(%): +10.64

```

Fig. 4.3: Salida por pantalla de la información del día 18 de marzo de 2016 para ACS.MC

El modelo se desplaza un día siendo ahora el día actual el 19 de marzo de 2016. En este momento, los 4 años anteriores comprenden aproximadamente entre el 26 de febrero de 2012 y el 26 de febrero de 2016. Con lo cual, ha entrado un día y ha salido otro en el modelo deslizante. Se actualizan los clusters y se le asigna al día actual, 19 de marzo de 2016, uno de ellos y su rentabilidad como predicción futura.

El modelo se mueve un día y se vuelve a repetir el mismo procedimiento, y así sucesivamente. En la figura 4.4 se muestra la predicción futura de la rentabilidad de cada uno de los días y su rentabilidad real, ya que disponemos de los datos de la rentabilidad futura.

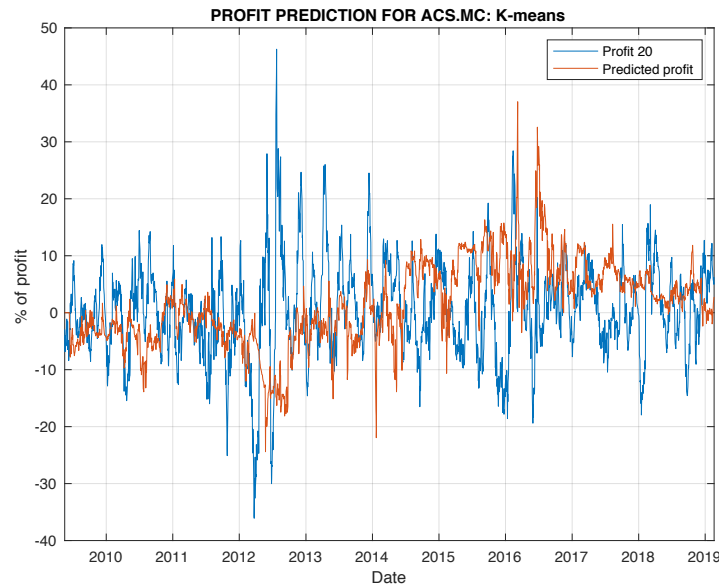


Fig. 4.4: Predicción de la rentabilidad futura para ACS.MC

Se puede observar la relación entre la rentabilidad predicha y la rentabilidad real en la figura 4.5. Cuanto más positiva sea esta correlación, más se parece la predicción de rentabilidad a la rentabilidad real, con lo cual, hay más posibilidades de acertar que de fallar si te basas en esta estrategia a la hora de invertir. En este caso, la correlación vale 0.00, con lo cual, no aportaría información relevante.

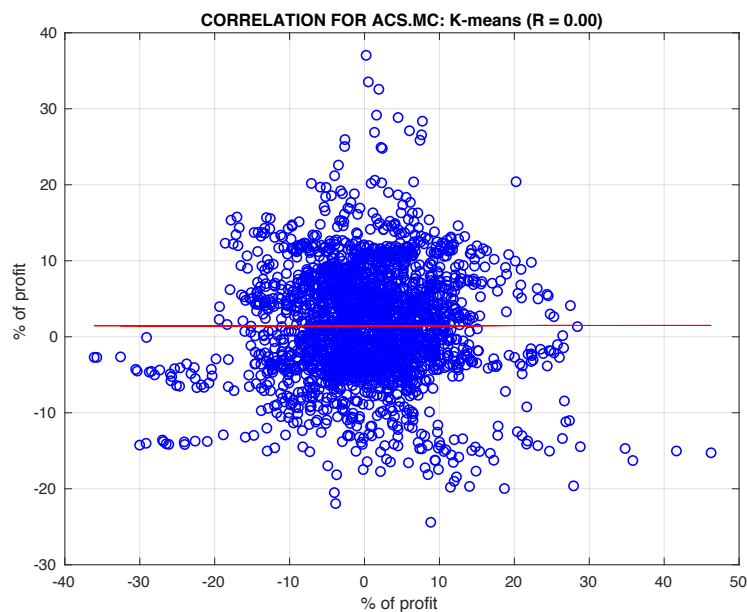


Fig. 4.5: Correlación entre rentabilidad predicha y rentabilidad real para ACS.MC

En la figura 4.6 se puede observar una correlación positiva de 0.12 en la acción DIA.MC, lo que indica que en esta acción se parece más la rentabilidad predicha a la rentabilidad futura real que en la acción ACS.MC.

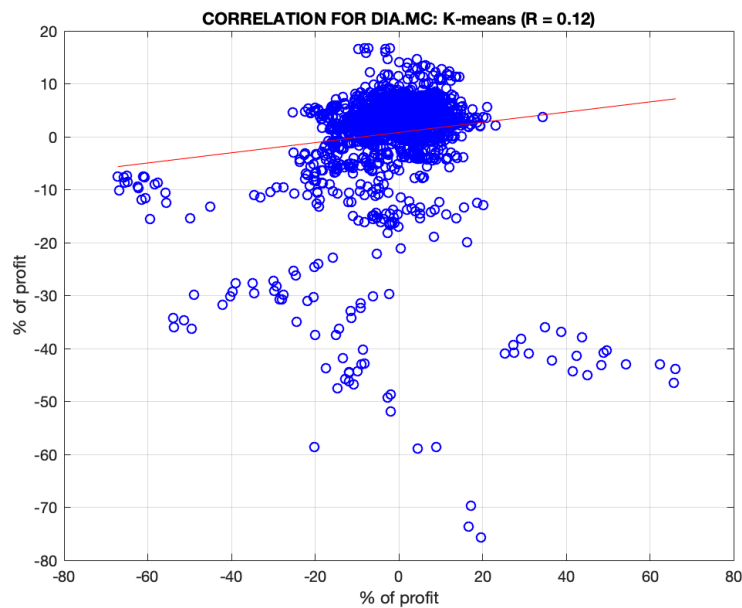


Fig. 4.6: Correlación entre rentabilidad predicha y rentabilidad real para DIA.MC

Los clusters, como se ha comentado, se actualizan día a día cuando se va moviendo la ventana deslizante. En la figura 4.7 se observa el movimiento de cada uno de los centroides en la sucesión de años.

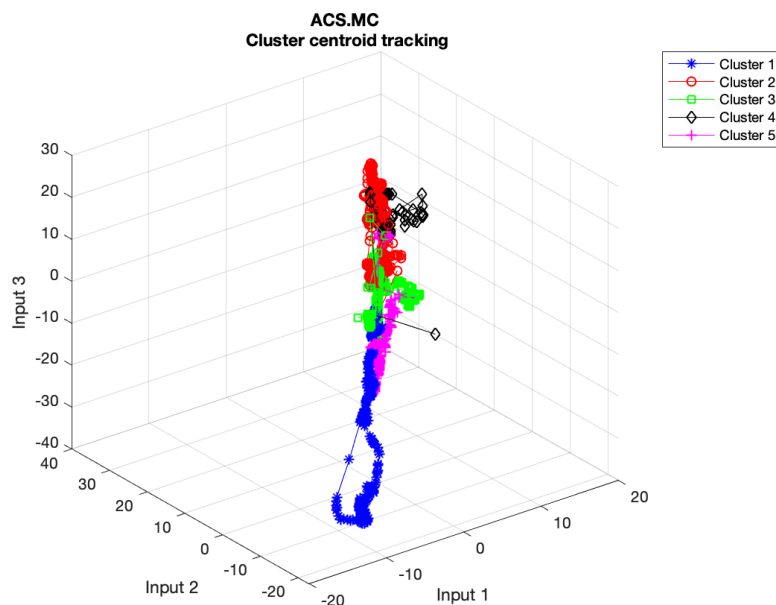


Fig. 4.7: Movimiento de los centroides para ACS.MC

La figura 4.8 representa el movimiento de las coordenadas u_1 , u_2 , u_3 , por separado, para cada uno de los 5 clusters.

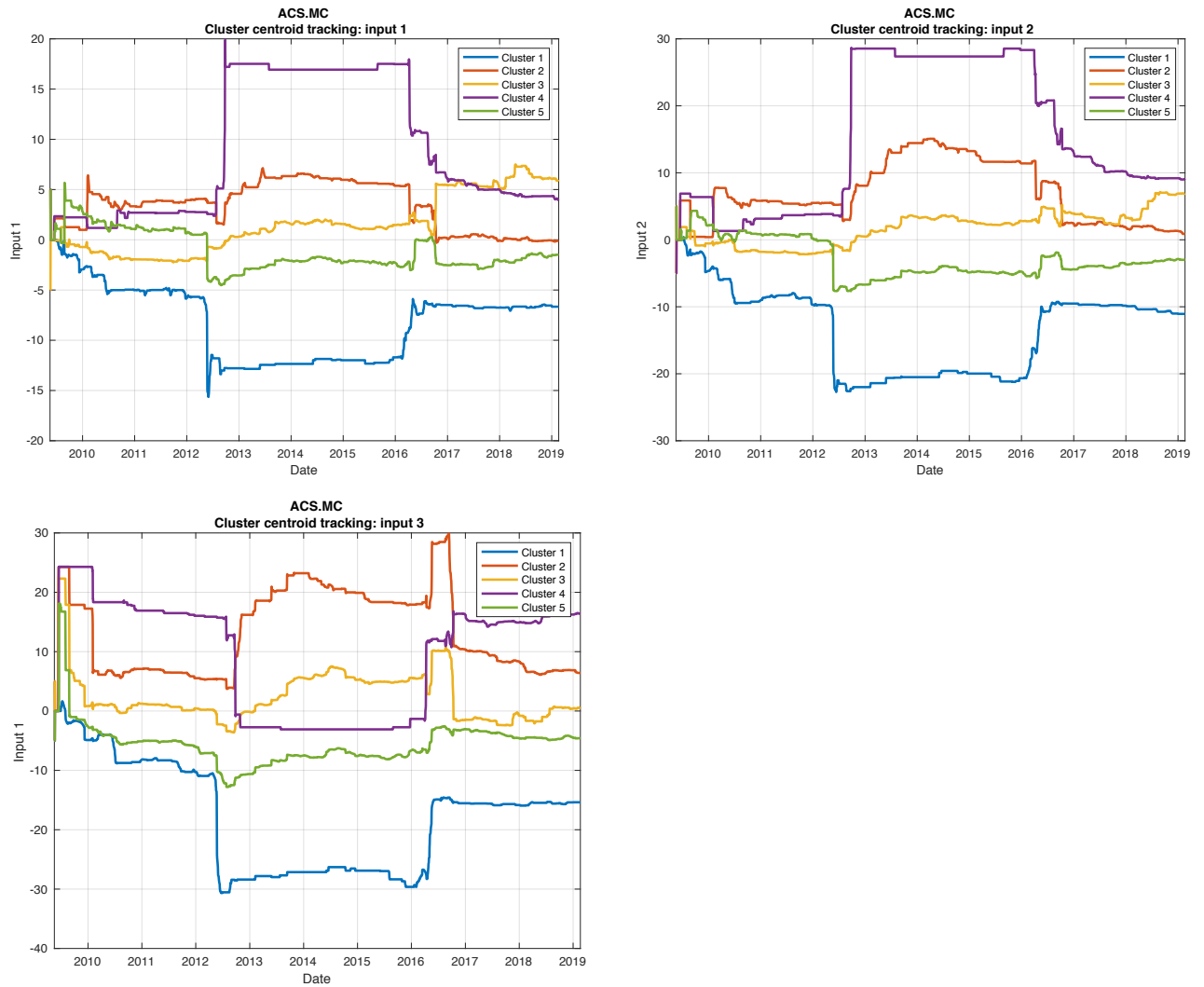


Fig. 4.8: Evolución temporal de las coordenadas u_1 , u_2 , u_3 para cada uno de las clusters para ACS.MC

En la figura 4.9 se muestra la evolución de las 3 entradas (u_1 , u_2 , u_3) para cada uno de los 5 clusters.

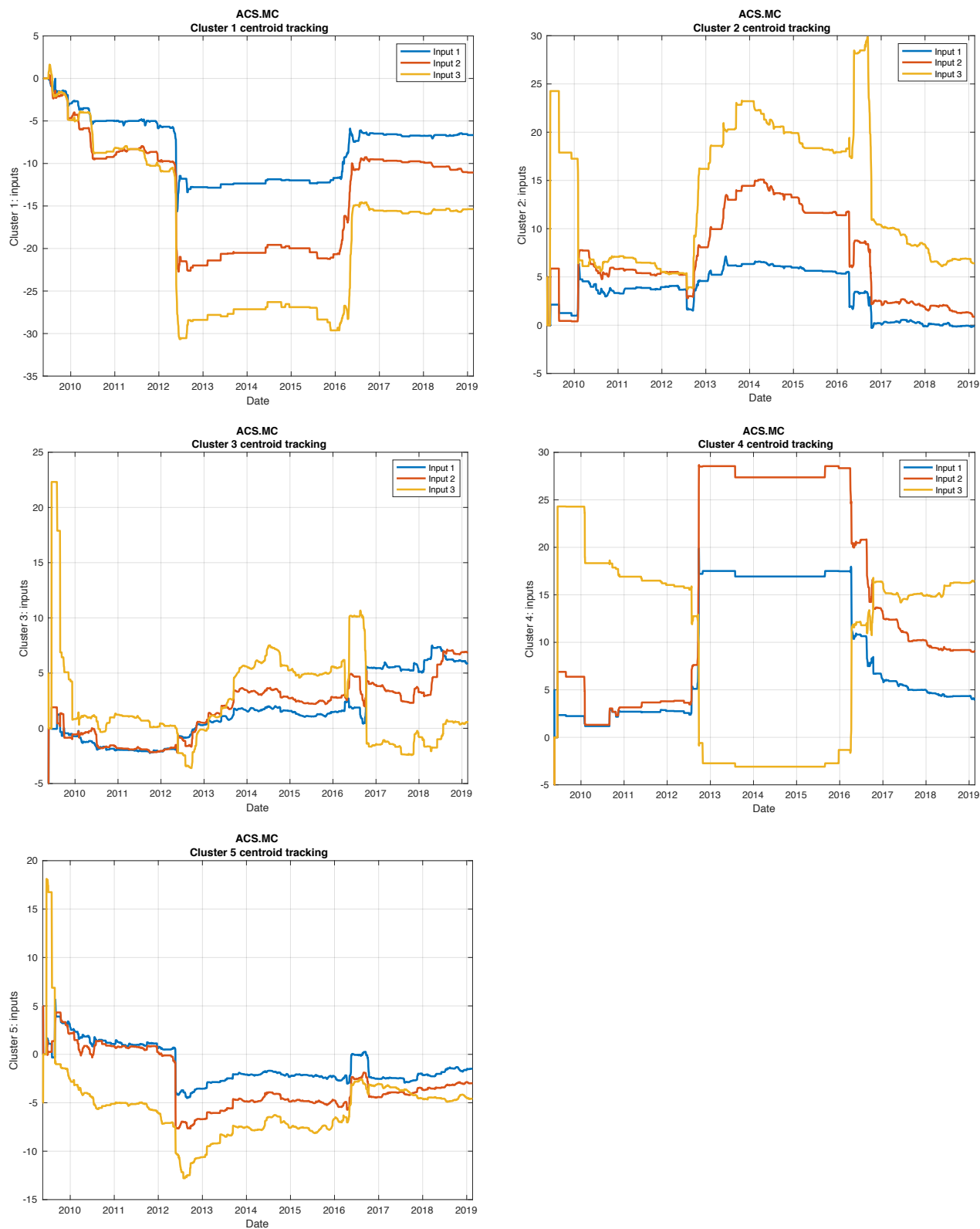


Fig. 4.9: Evolución temporal de las tres entradas para cada uno de los 5 clusters para ACS.MC

Por último, la figura 4.10 representa el movimiento de las rentabilidades asignadas a cada cluster y la figura 4.11 el movimiento de las rentabilidades promediadas. En este último se da más peso a aquellos puntos más próximos al centroide correspondiente.

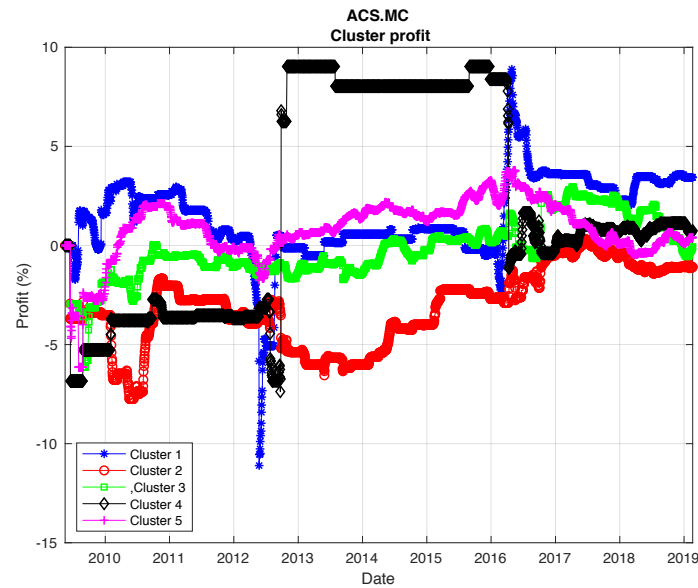


Fig. 4.10: Evolución de las rentabilidades de cada cluster para ACS.MC

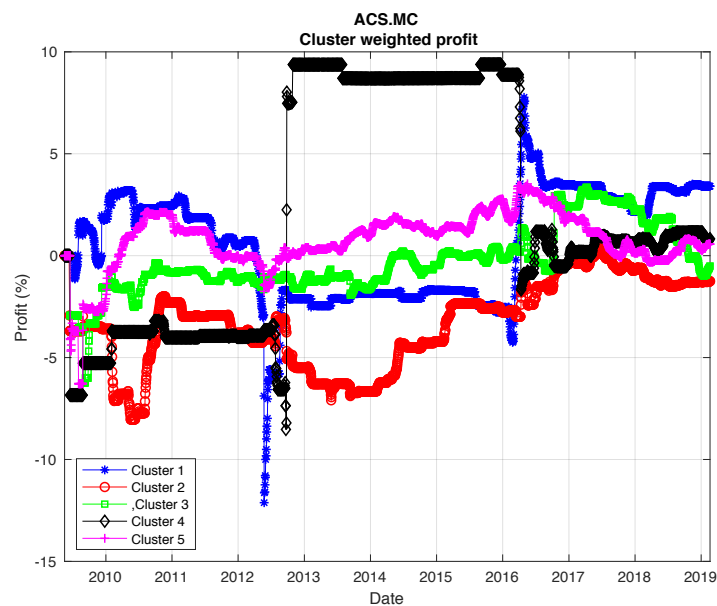


Fig. 4.11: Evolución de las rentabilidades promediadas de cada cluster para ACS.MC

Una vez calculadas las rentabilidades predichas de cada día, se genera la señal de compra-venta representada en la figura 4.12. Como se ha explicado en el capítulo anterior, existen diferentes técnicas de estrategia de inversión.

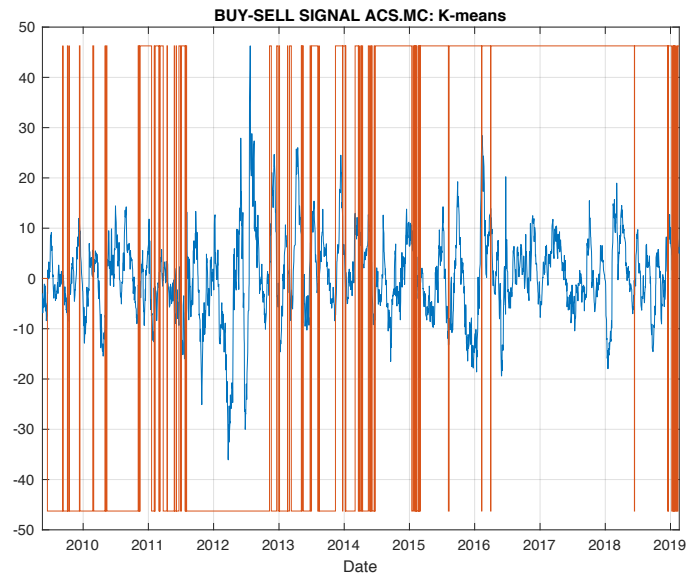


Fig. 4.12: Señal de compra-venta para ACS.MC

La gráfica de compra-venta indica, como su nombre indica, cuando se debe comprar y vender una determinada acción. Puede valer -1, 1 y 0. Cuando vale -1 indica que se tiene que venderla acción si has comprado previamente, si vale 0 que no se realice ninguna transacción y, si vale 1, indica que, si no has comprado previamente, compres en ese instante.

En el caso de la figura 4.12 vale 40 y -40 en lugar de 1 y -1 porque la gráfica se encuentra escalada con respecto al máximo.

En la figura 4.13 se representa aquella estrategia de inversión que utiliza la información proporcionada por la señal de compra-venta para comprar y vender. En la gráfica aparecen las rentabilidades acumuladas.

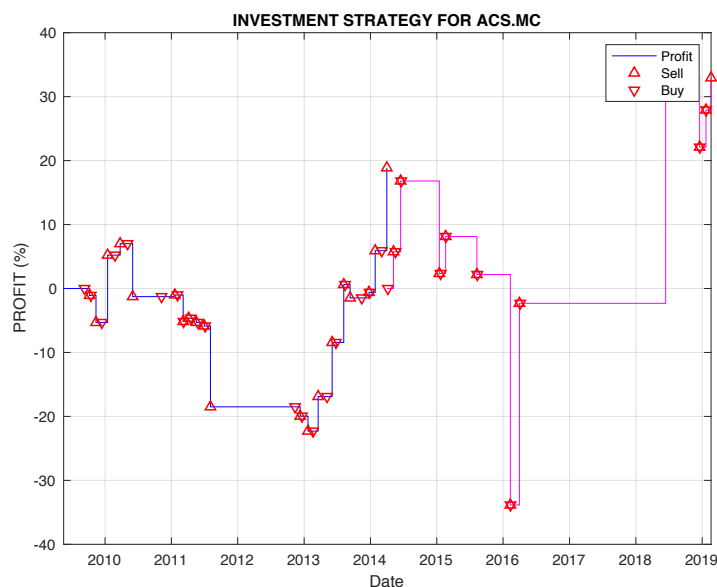


Fig. 4.13: Estrategia de inversión para ACS.MC

La figura 4.14 representa la estrategia complementaria a la señal de compra-venta. Como se puede observar, compra y vende al contrario que la gráfica anterior.

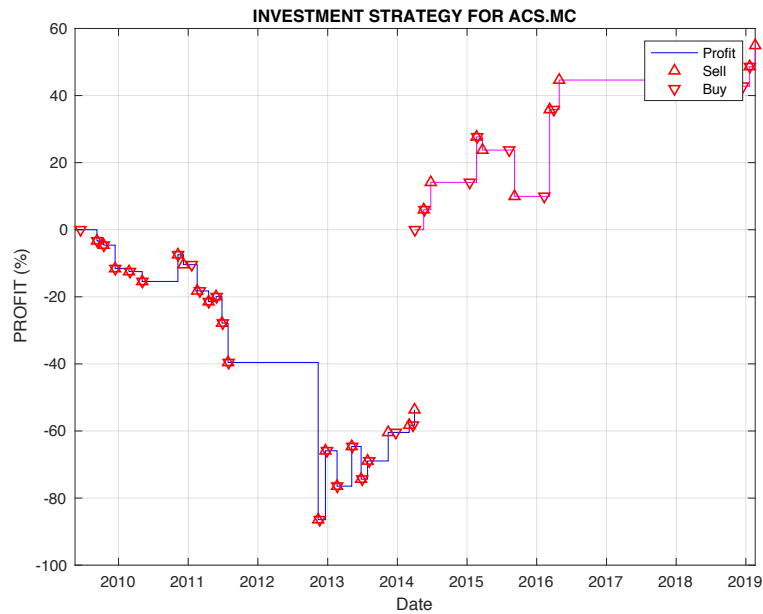


Fig. 4.14: Estrategia de inversión complementaria para ACS.MC

Por último, la figura 4.15 representa la estrategia de inversión en la cual se compra cuando la rentabilidad predicha está por encima de un umbral y se vende después de la longitud de la ventana futura, en este caso 20 días.

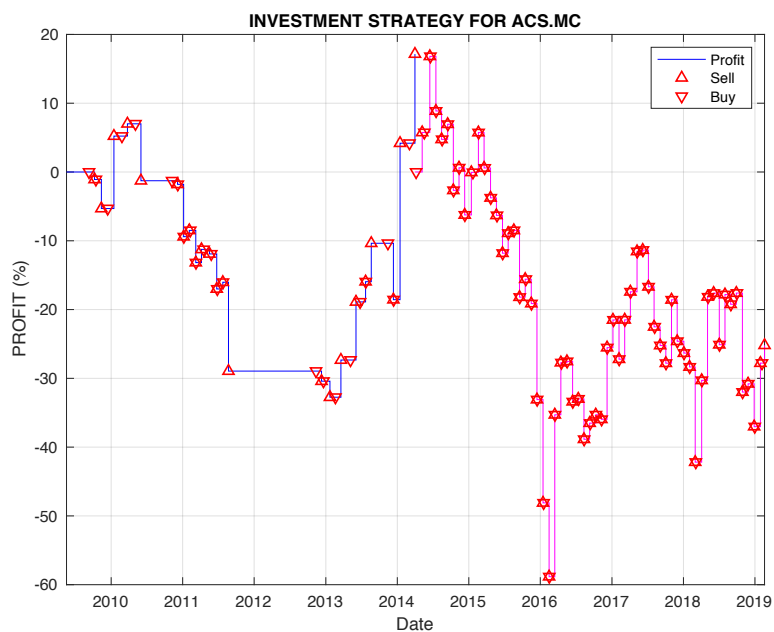


Fig. 4.15: Estrategia de inversión fija para ACS.MC

En la figura 4.16 se analizan los datos de la estrategia de inversión durante los 10 años de la figura 4.13, es decir, aquella que sigue la señal de compra-venta para realizar las transacciones correspondientes.

Los datos obtenidos para cada uno de los tramos, optimización y test, son:

- Fecha inicial, final y número de años que abarca cada uno de los tramos, tanto optimización como test.
- Rentabilidad total acumulada.
- Rentabilidad compuesta, es decir, aquella que se obtiene cuando se va reinvertiendo lo que se va obteniendo.
- Rentabilidad pasiva, es decir, aquella que se obtiene cuando se invierte al principio y se recupera al final.
- Número de operaciones de compra-venta realizadas en total en cada intervalo, cuántas de éstas generan rentabilidad positiva, cuántas generan rentabilidad negativa y número de operaciones de compra-venta realizadas por año.
- Máxima y mínima rentabilidad acumulada.
- Máxima racha de rentabilidad positiva y mínima racha de rentabilidad negativa.
- Rentabilidad media que se obtiene en cada operación de compra-venta.
- Máxima rentabilidad y mínima en las operaciones de compra-venta.
- Duración máxima, mínima y media de las operaciones de compra-venta.
- Tiempo total con operaciones de compra-venta activas.
- Coste de cada operación.
- Aparecen las fechas de compra y venta de cada una de las operaciones realizadas durante el intervalo y la rentabilidad obtenida en dicha operación.

Esta tabla no aporta información relevante en este caso debido a que no se han utilizado los metaparámetros óptimos como se hará en el capítulo 6.

El tramo de optimización está formado por el 50% de los datos, previamente determinado y correspondiente los 50% restantes al test.

El entrenamiento y validación se realizan del día 18 de mayo de 2009 al 14 de abril del 2014, mientras que el test se realiza del 2 de abril de 2014 al 19 de febrero de 2019.

A parte de la rentabilidad, es importante tener en cuenta otros índices a la hora de invertir que se analizarán a continuación.

Teniendo en cuenta el número de operaciones negativas y positivas realizadas durante estos 4 años, en la parte correspondiente a entrenamiento y validación son 50% y 50% respectivamente mientras que en el test estos valores son 63.6% y 36.4% indicando que existen más operaciones positivas que negativas lo cual podría ser beneficioso a la hora de invertir. En realidad, este porcentaje puede no ser determinante sin tener en cuenta cada operación debido a que puede que las operaciones positivas aporten una

rentabilidad pequeña mientras que las negativas, aunque se produzcan en menos ocasiones, afecten más.

En este caso, en general, las operaciones positivas tienen más peso que las negativas.

La rentabilidad máxima y mínima en el entrenamiento y validación son 18.9% y -22.3% mientras que en el test son 33.2% y -33.9%, difiriendo de lo indicado en el entrenamiento.

Es importante tener en cuenta la racha negativa de rentabilidad. Teniendo en cuenta que, durante el test, la rentabilidad final es 32.9 % y que la mínima racha negativa es -50.7%, se podría invertir en un momento inadecuado llegando a perder un 50.7% de la cantidad invertida.

La figura 4.16 también aporta información sobre la rentabilidad obtenida en cada una de las operaciones realizadas y, se puede observar, que esta rentabilidad no se adquiere de forma continua.

INVESTMENT STRATEGY RESULTS FOR OPTIMIZATION

INITIAL DATE: 18-May-2009
 FINAL DATE: 01-Apr-2014
 NUMBER OF YEARS: 4.9
 TOTAL PROFIT (%): +18.9
 PROFIT/YEAR (%): +3.9
 COMPOUND PROFIT (%): +16.2
 PASSIVE PROFIT (%): -20.4
 NUMBER OF OPERATIONS: 20
 NUMBER OF OPERATIONS/YEAR: 4.1
 % OF POSITIVE OPERATIONS: 50.0
 % OF NEGATIVE OPERATIONS: 50.0
 MAXIMUM PROFIT (%): +18.9
 MINIMUM PROFIT (%): -22.3
 MAXIMUM PROFIT STREAK (%): +41.2
 MINIMUM PROFIT STREAK (%): -29.3
 AVERAGE PROFIT/OPERATION (%): +0.9
 MAXIMUM PROFIT/OPERATION (%): +13.0
 MINIMUM PROFIT/OPERATION (%): -12.6
 AVERAGE OPERATION DURATION (days): 22.9
 MAXIMUM OPERATION DURATION (days): 51
 MINIMUM OPERATION DURATION (days): 20
 ACTIVE OPERATION TIME (%): 18.3
 COST OF BUY-SELL OPERATION (%): 1.0
 OPERATIONS:

Buy date	Sell date	Days	Profit(%)
10-Sep-2009	08-Oct-2009	20	-1.1
15-Oct-2009	12-Nov-2009	20	-4.2
14-Dec-2009	15-Jan-2010	20	+10.5
25-Feb-2010	25-Mar-2010	20	+1.8
05-May-2010	02-Jun-2010	20	-8.3
08-Nov-2010	20-Jan-2011	51	+0.3
04-Feb-2011	07-Mar-2011	21	-4.2
09-Mar-2011	06-Apr-2011	20	+0.5
18-Apr-2011	18-May-2011	20	-0.7
27-May-2011	01-Jul-2011	25	-0.6
06-Jul-2011	04-Aug-2011	21	-12.6
12-Nov-2012	10-Dec-2012	20	-1.5
20-Dec-2012	22-Jan-2013	20	-2.3
19-Feb-2013	19-Mar-2013	20	+5.4
07-May-2013	04-Jun-2013	20	+8.4
25-Jun-2013	07-Aug-2013	31	+9.1
14-Aug-2013	11-Sep-2013	20	-2.1
14-Nov-2013	24-Dec-2013	28	+0.9
27-Dec-2013	27-Jan-2014	20	+6.5
04-Mar-2014	01-Apr-2014	20	+13.0

INVESTMENT STRATEGY FOR TEST

INITIAL DATE: 02-Apr-2014
 FINAL DATE: 19-Feb-2019
 NUMBER OF YEARS: 4.9
 TOTAL PROFIT (%): +32.9
 PROFIT/YEAR (%): +6.7
 COMPOUND PROFIT (%): +12.6
 PASSIVE PROFIT (%): +28.1
 NUMBER OF OPERATIONS: 11
 NUMBER OF OPERATIONS/YEAR: 2.2
 % OF POSITIVE OPERATIONS: 63.6
 % OF NEGATIVE OPERATIONS: 36.4
 MAXIMUM PROFIT (%): +33.2
 MINIMUM PROFIT (%): -33.9
 MAXIMUM PROFIT STREAK (%): +67.0
 MINIMUM PROFIT STREAK (%): -50.7
 AVERAGE PROFIT/OPERATION (%): +3.0
 MAXIMUM PROFIT/OPERATION (%): +35.5
 MINIMUM PROFIT/OPERATION (%): -36.0
 AVERAGE OPERATION DURATION (days): 111.1
 MAXIMUM OPERATION DURATION (days): 563
 MINIMUM OPERATION DURATION (days): 20
 ACTIVE OPERATION TIME (%): 88.9
 COST OF BUY-SELL OPERATION (%): 1.0
 OPERATIONS:

Buy date	Sell date	Days	Profit(%)
07-Apr-2014	08-May-2014	20	+5.7
19-May-2014	16-Jun-2014	20	+11.1
19-Jun-2014	15-Jan-2015	147	-14.5
22-Jan-2015	19-Feb-2015	20	+5.8
20-Feb-2015	10-Aug-2015	118	-5.9
12-Aug-2015	10-Feb-2016	128	-36.0
11-Feb-2016	31-Mar-2016	33	+31.5
04-Apr-2016	14-Jun-2018	563	+35.5
15-Jun-2018	18-Dec-2018	132	-11.1
19-Dec-2018	22-Jan-2019	21	+5.8
23-Jan-2019	19-Feb-2019	20	+5.0

Fig. 4.16: Salida por pantalla de la información de la estrategia de inversión para ACS.MC

5. GMM

Como se ha explicado anteriormente, el método GMM utiliza un modelo de suma de distribuciones gaussianas n-dimensionales sobre los datos. Las gaussianas se ajustan en las zonas donde hay más acumulación de puntos. Cada punto pertenecerá a la gaussiana a la que tenga mayor porcentaje de pertenencia.

Con los puntos de la figura 3.21 se forman las 5 gaussianas del modelo. En la tabla 5.1 se definen los centroides, es decir, las medias de cada una de las gaussianas formadas con los puntos correspondientes a los 4 años.

CENTROIDES			
	Componente x	Componente y	Componente z
Centroide 1	-0.03	+0.55	-0.05
Centroide 2	+2.11	+2.25	+16.26
Centroide 3	+0.05	-0.29	-0.19
Centroide 4	-1.28	-0.74	-1.16
Centroide 5	+20.13	+18.40	-3.80

Tabla 5.1: Coordenadas de las medias de las gaussianas para ACS.MC en el día 18 de marzo de 2016

Una vez se dispone de las gaussianas, se calcula la rentabilidad futura de cada una de ellas como la media de las rentabilidades futuras de los puntos que pertenecen a cada una. Además, al igual que en el método K-means se aportan otros datos de cada una de las gaussianas que serán definidos en la tabla 5.2.

Tanto la población, dispersión de los puntos dentro de cada cluster, rentabilidad del cluster, volatilidad interna y volatilidad externa se calculan de la misma manera que en el método K-means. Para rentabilidad promediada del cluster no se utiliza la participación del punto en el cluster si no que se calcula directamente con la probabilidad de la gaussiana a la que pertenece.

GAUSSIAN INFO						
PPL (%)	DSP (%)	PRF (%)	WPR (%)	IVL (%)	FVL (%)	KPI
0.1	0.0	+18.8	+18.8	6.3	0.0	0.00
0.2	0.1	-9.5	-9.5	1.4	0.9	0.00
70.9	30.5	+1.3	+1.4	3.2	9.2	1.13
27.4	42.6	-0.1	0.0	4.1	12.5	0.00
1.4	26.9	+11.6	+11.8	4.7	4.8	0.62

Tabla 5.2: Información de cada gaussiana para ACS.MC en el día 18 de marzo de 2016

El día actual, 18 de marzo de 2016, dispone de unas entradas mostradas en la figura 5.1. En éstas se muestra la rentabilidad de los últimos 10,20 y 40 días. Con lo cual este punto corresponde a un punto en el espacio 3D que pertenecerá con más probabilidad a una de las gaussianas creadas con los 4 años anteriores.

```

INPUTS FOR PREDICTION MODEL:
SHORT PROFIT FOR THE LAST 10 DAYS(%):  3.87
MEDIUM PROFIT FOR THE LAST 20 DAYS(%): 24.36
LONG PROFIT FOR THE LAST 40 DAYS(%):  16.63

```

Fig. 5.1: Salida por pantalla de los datos de las señales de entrada para ACS.MC

Como se muestra en la figura 5.2, el día 18 de marzo de 2016 con las coordenadas [+3.87 +24.36 +16.63] tiene un 99.5% de probabilidad de pertenecer a la gaussiana 4. Por tanto, al ser ésta la gaussiana con mayor probabilidad, se dice que pertenece a la gaussiana 4. Además, se le asigna la rentabilidad de esta gaussiana, siendo ésta -0.14%.

En este caso, el current weighted profit representa la media de las probabilidades de pertenencia a cada gaussiana, por eso no coincide con la rentabilidad promediada de la gaussiana 4, ya que se calculan de maneras diferentes.

En comparación con el método K-means, se observa que, debido a que los clusters en K-means se forma de acuerdo a la mínima distancia euclídea, estos tienen coordenadas distintas y, por lo tanto, no corresponden con las medias de las gaussianas ni con las rentabilidades asignadas a cada cluster.

```

CURRENT PARAMETER VECTOR INFO
PARAMETER VECTOR(%): [ +3.87 +24.36 +16.63]
MEMBERSHIP(%): [ 0.0  0.0  0.5 99.5  0.0]
CURRENT CLUSTER: 4
CURRENT MEAN PROFIT(%):  -0.14
CURRENT WEIGHTED PROFIT(%):  -0.13
CURRENT SCALED MEAN PROFIT(%):  -0.57
CURRENT SCALED WEIGHTED PROFIT(%):  -0.54

```

Fig. 5.2: Salida por pantalla de la información del día 18 de marzo de 2016 para ACS.MC

Como se ha explicado en el capítulo 3, una vez se sabe a qué gaussiana pertenece el día actual, el modelo se mueve un día y se realiza el mismo procedimiento para calcular la rentabilidad predicha de cada punto. Así, se forma la curva naranja de la figura 5.3, siendo la curva azul, la rentabilidad real de los puntos.

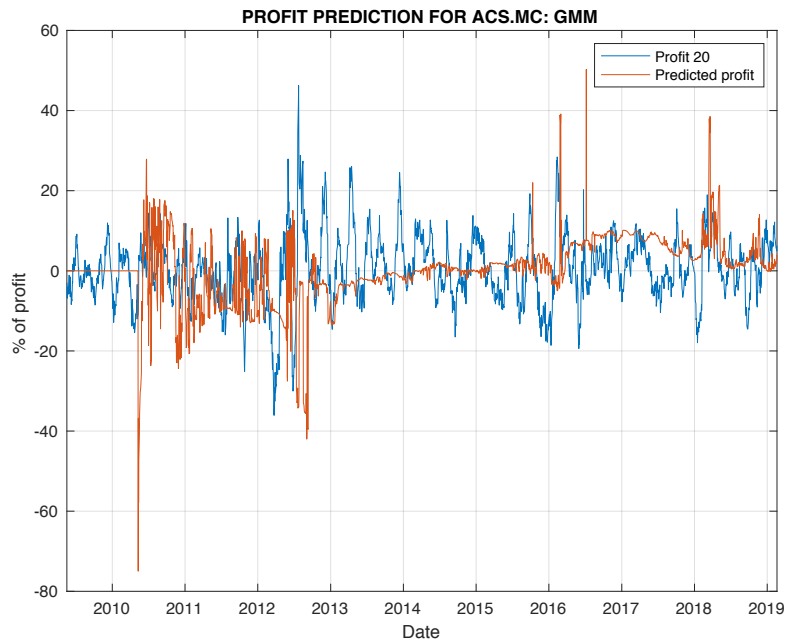


Fig. 5.3: Predicción de la rentabilidad futura para ACS.MC

Seguidamente, se calcula la correlación entre ambas curvas en la figura 5.4, siendo ésta de 0.0, con lo cual, al igual que utilizando la técnica del K-means, no aporta información relevante.

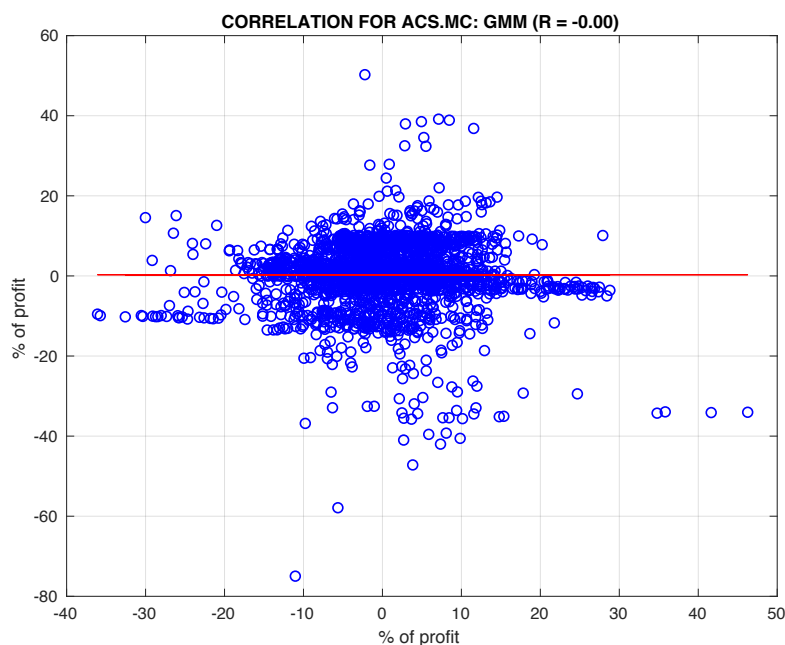


Fig. 5.4: Correlación entre rentabilidad predicha y rentabilidad real para ACS.MC

De la misma manera que se hizo en el K-means, se ha calculado la correlación entre la rentabilidad predicha y la rentabilidad real para la acción DIA.MC y, en contraste con la técnica de K-means, utilizando GMM, la correlación es -0.01 en lugar de 0.04.

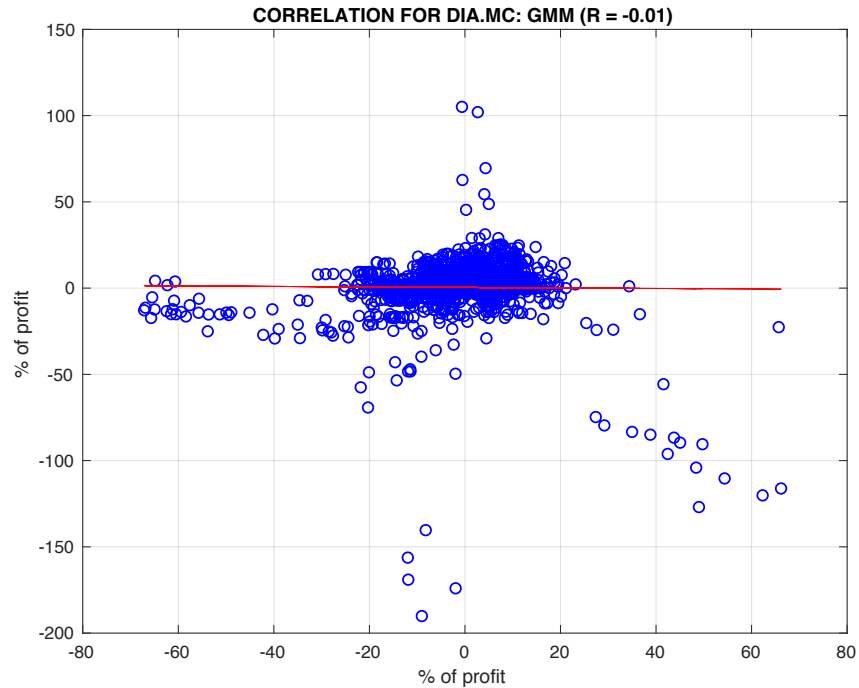


Fig. 5.5: Correlación entre rentabilidad predicha y rentabilidad real para DIA.MC

Cada media de gaussiana dispone de 3 coordenadas que van cambiando a medida que va entrando y saliendo un punto del modelo.

En la figura 5.6 se representa el movimiento de las medias de las 5 gaussianas en el espacio 3D. En la figura 5.7, se muestra la evolución temporal de las coordenadas u_1 , u_2 , y u_3 , por separado, de cada una de las gaussianas. Y, finalmente, en la figura 5.8 se muestra en cada gráfica la evolución temporal de las 3 coordenadas de cada gaussiana.

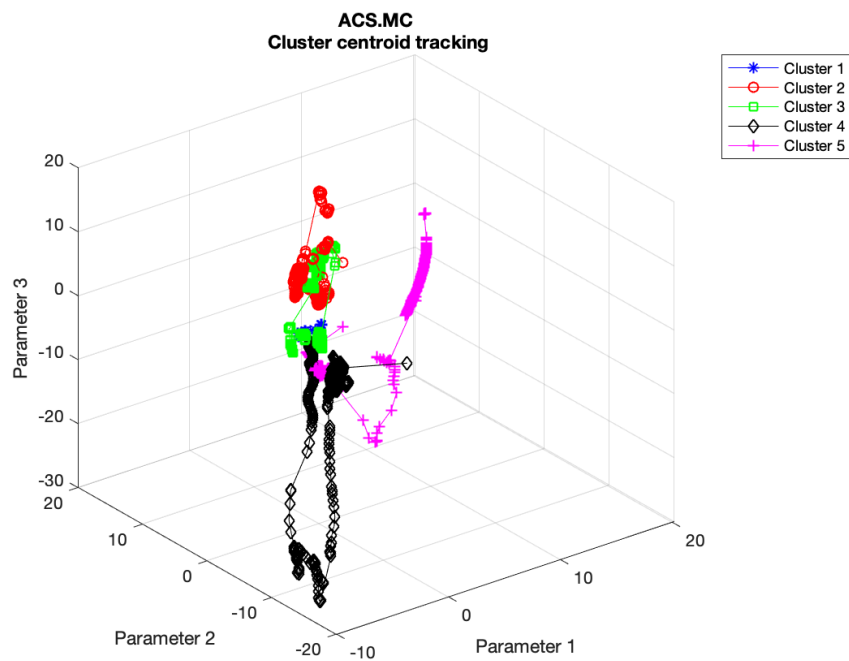


Fig. 5.6: Movimiento de las medias de las gaussianas para ACS.MC

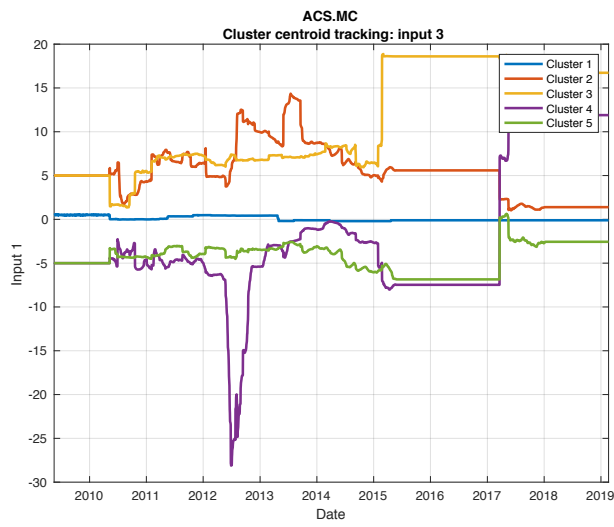
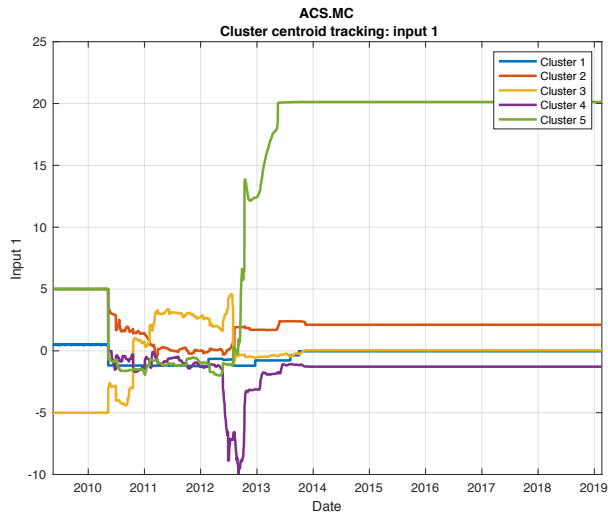


Fig. 5.7: Evolución temporal de las coordenadas u_1 , u_2 , u_3 para cada una de las medias de las gaussianas para ACS.MC

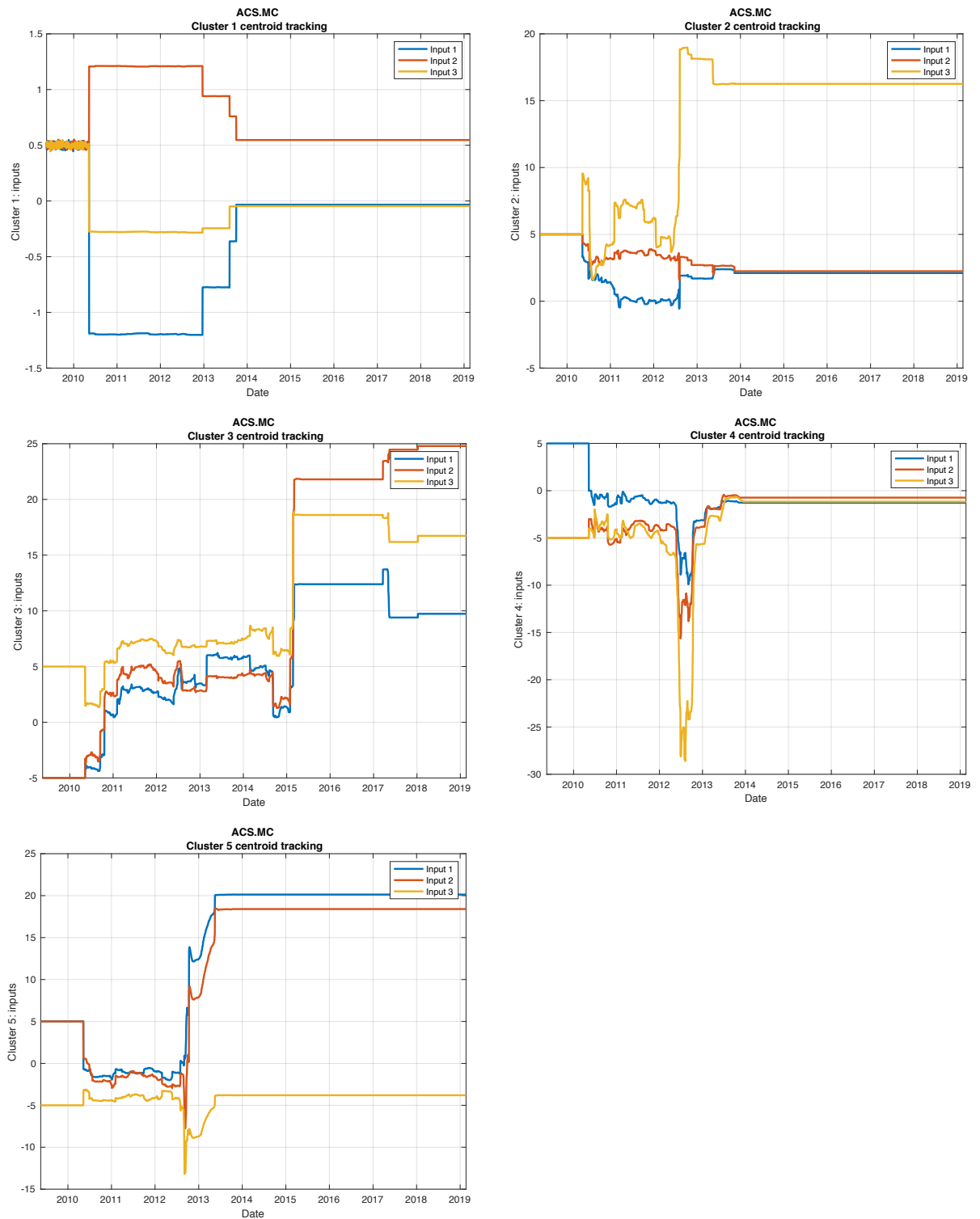


Fig. 5.8: Evolución temporal de las tres entradas para cada uno de las 5 gaussianas para ACS.MC

En las figuras 5.9 y 5.10 se representa la evolución temporal de la rentabilidad asignada a cada gaussiana y la rentabilidad promediada, respectivamente. Como se ha explicado anteriormente, ésta va cambiando según los puntos que van entrando y saliendo del intervalo de 4 años.

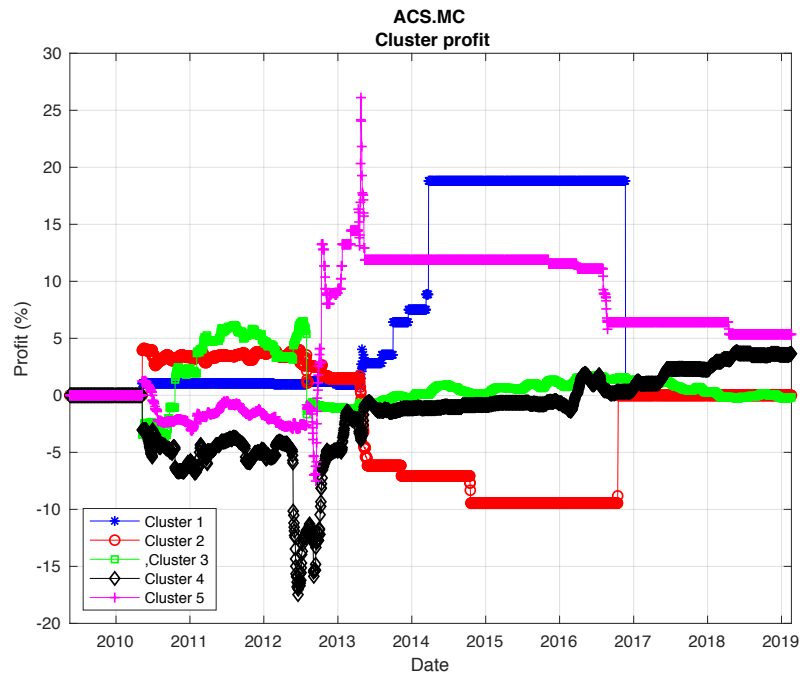


Fig. 5.9: Evolución de las rentabilidades de cada gaussiana para ACS.MC

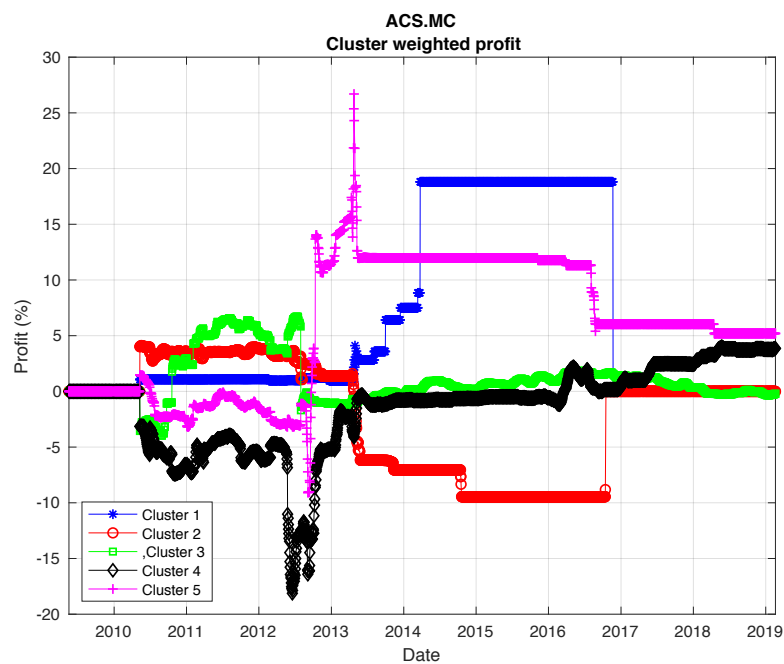


Fig. 5.10: Evolución de las rentabilidades promediadas de cada gaussiana para ACS.MC

Tras el cálculo de las rentabilidades predichas de cada día, se genera la señal de compra-venta de la figura 5.11, que sigue las mismas normas que en el método K-means. Cuando vale 1 indica que se compre, 0 que no se realice ninguna operación y -1 que se venda la acción. En este caso es un valor entre 40 y 50 ya que está escalado con el máximo.

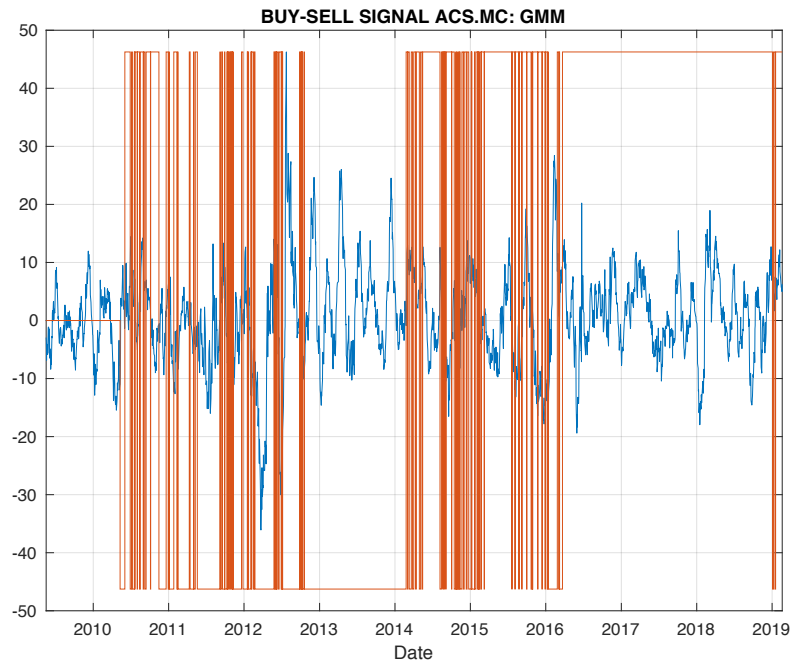


Fig. 5.11: Estrategia de inversión para ACS.MC

Una vez calculada la señal de compra-venta, se genera la estrategia de inversión de la misma manera que se generó en el método K-means. Así, se puede representar la estrategia de inversión que sigue la señal de compra-venta en la figura 5.12, aquella que sigue la señal complementaria a la de compra-venta en la figura 5.13 y aquella estrategia de inversión en la cual se compra cuando la rentabilidad predicha está por encima de un umbral y se vende después de la longitud de la ventana futura, en este caso 20 días.

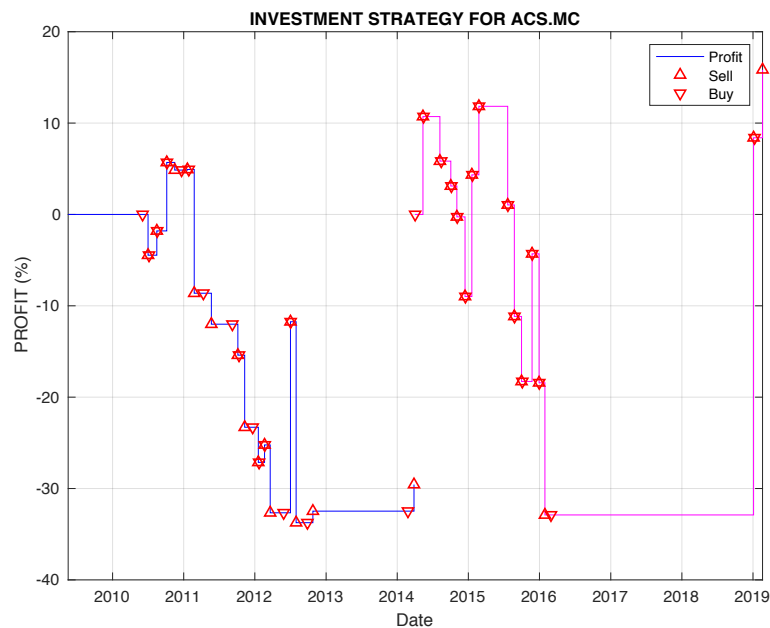


Fig. 5.12: Estrategia de inversión para ACS.MC

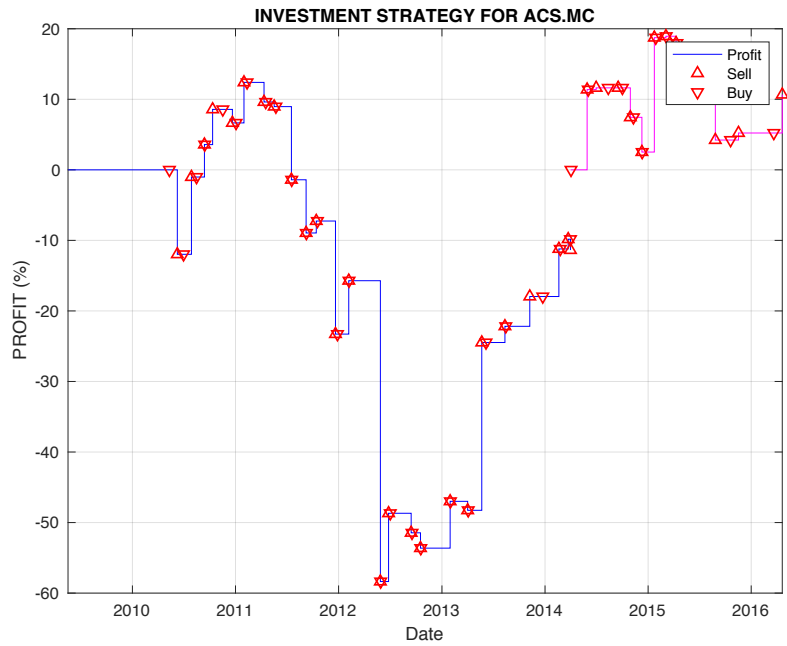


Fig. 5.13: Estrategia de inversión complementaria para ACS.MC

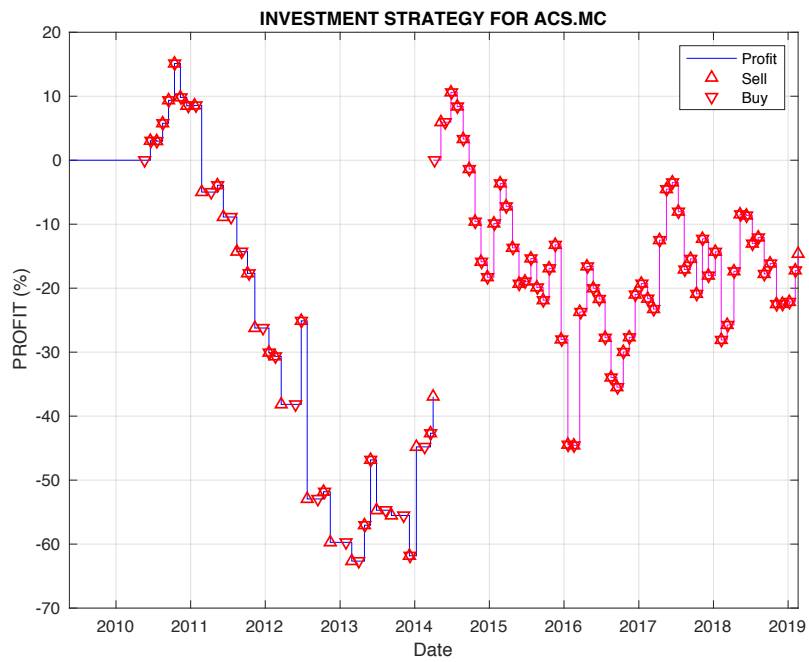


Fig. 5.14: Estrategia de inversión fija para ACS.MC

Los datos de la figura 5.15, al igual que en el caso del K-means, no son relevantes debido a que no se han utilizados los metaparámetros óptimos.

INVESTMENT STRATEGY ANALYSIS				
INVESTMENT STRATEGY PARAMETERS:				
LONG WINDOW: 40				
MEDIUM WINDOW: 20				
SHORT WINDOW: 10				
HIGH THRESHOLD (%): +0				
LOW THRESHOLD (%): +0				
INVESTMENT STRATEGY RESULTS FOR OPTIMIZATION				
INITIAL DATE: 18-May-2009				
FINAL DATE: 01-Apr-2014				
NUMBER OF YEARS: 4.9				
TOTAL PROFIT (%): -29.6				
PROFIT/YEAR (%): -6.1				
COMPOUND PROFIT (%): -30.7				
PASSIVE PROFIT (%): -6.6				
NUMBER OF OPERATIONS: 16				
NUMBER OF OPERATIONS/YEAR: 3.3				
% OF POSITIVE OPERATIONS: 43.8				
% OF NEGATIVE OPERATIONS: 56.2				
MAXIMUM PROFIT (%): +5.7				
MINIMUM PROFIT (%): -33.7				
MAXIMUM PROFIT STREAK (%): +10.2				
MINIMUM PROFIT STREAK (%): -39.4				
AVERAGE PROFIT/OPERATION (%): -1.8				
MAXIMUM PROFIT/OPERATION (%): +20.9				
MINIMUM PROFIT/OPERATION (%): -22.0				
AVERAGE OPERATION DURATION (days): 22.9				
MAXIMUM OPERATION DURATION (days): 35				
MINIMUM OPERATION DURATION (days): 20				
ACTIVE OPERATION TIME (%): 14.7				
COST OF BUY-SELL OPERATION (%): 1.0				
OPERATIONS:				
Buy date	Sell date	Days	Profit(%)	
04-Jun-2010	02-Jul-2010	20	-4.5	
07-Jul-2010	16-Aug-2010	28	+2.7	
18-Aug-2010	06-Oct-2010	35	+7.5	
07-Oct-2010	16-Nov-2010	28	-0.8	
21-Dec-2010	20-Jan-2011	20	+0.1	
27-Jan-2011	24-Feb-2011	20	-13.5	
12-Apr-2011	23-May-2011	27	-3.4	
08-Sep-2011	06-Oct-2011	20	-3.4	
12-Oct-2011	10-Nov-2011	21	-7.9	
21-Dec-2011	19-Jan-2012	20	-3.8	
23-Jan-2012	20-Feb-2012	20	+1.9	
21-Feb-2012	20-Mar-2012	20	-7.4	
28-May-2012	02-Jul-2012	25	+20.9	
03-Jul-2012	31-Jul-2012	20	-22.0	
27-Sep-2012	25-Oct-2012	20	+1.3	
25-Feb-2014	28-Mar-2014	23	+2.9	
INVESTMENT STRATEGY FOR TEST				
INITIAL DATE: 02-Apr-2014				
FINAL DATE: 19-Feb-2019				
NUMBER OF YEARS: 4.9				
TOTAL PROFIT (%): +15.8				
PROFIT/YEAR (%): +3.2				
COMPOUND PROFIT (%): +1.8				
PASSIVE PROFIT (%): +30.7				
NUMBER OF OPERATIONS: 15				
NUMBER OF OPERATIONS/YEAR: 3.1				
% OF POSITIVE OPERATIONS: 40.0				
% OF NEGATIVE OPERATIONS: 60.0				
MAXIMUM PROFIT (%): +15.8				
MINIMUM PROFIT (%): -32.9				
MAXIMUM PROFIT STREAK (%): +48.7				
MINIMUM PROFIT STREAK (%): -44.7				
AVERAGE PROFIT/OPERATION (%): +1.1				
MAXIMUM PROFIT/OPERATION (%): +41.3				
MINIMUM PROFIT/OPERATION (%): -14.5				
AVERAGE OPERATION DURATION (days): 80.1				
MAXIMUM OPERATION DURATION (days): 729				
MINIMUM OPERATION DURATION (days): 20				
ACTIVE OPERATION TIME (%): 51.3				
COST OF BUY-SELL OPERATION (%): 1.0				
OPERATIONS:				
Buy date	Sell date	Days	Profit(%)	
03-Apr-2014	13-May-2014	25	+10.7	
16-May-2014	08-Aug-2014	60	-4.9	
15-Aug-2014	03-Oct-2014	35	-2.7	
06-Oct-2014	03-Nov-2014	20	-3.4	
05-Nov-2014	15-Dec-2014	28	-8.7	
17-Dec-2014	19-Jan-2015	20	+13.3	
22-Jan-2015	24-Feb-2015	23	+7.5	
25-Feb-2015	22-Jul-2015	102	-10.8	
23-Jul-2015	25-Aug-2015	23	-12.2	
26-Aug-2015	01-Oct-2015	26	-7.1	
05-Oct-2015	24-Nov-2015	36	+14.0	
25-Nov-2015	30-Dec-2015	24	-14.1	
31-Dec-2015	29-Jan-2016	20	-14.5	
29-Feb-2016	04-Jan-2019	729	+41.3	
09-Jan-2019	19-Feb-2019	30	+7.4	

Fig. 5.15: Salida por pantalla de la información de la estrategia de inversión para ACS.MC

En este caso se observa que la rentabilidad final en el periodo de optimización es -29.6% y en el test +15.8% lo que aportaría beneficio al final de las operaciones. La mínima racha negativa de rentabilidad es -39.4% en optimización y -44.7% en el test.

Teniendo en cuenta esto y la rentabilidad por operaciones se puede observar que esta no se adquiere de manera continua en ninguno de los intervalos pudiendo esto perjudicar si se invierte en un momento inadecuado.

El número de operaciones positivas y negativas en ambos intervalos es muy parecido siendo ésta de 43.8%- 56.2% y de 40% - 60% para la optimización y test respectivamente. Estos datos contrastan con lo obtenido con el método K-means en el cual el número de operaciones negativas era mayor que el número de operaciones positivas. Por tanto, en el caso del GMM, habría que tener en cuenta lo contrario, que las operaciones positivas podrían aportar una rentabilidad pequeña mientras que las negativas podrían afectar más.

6. Análisis de resultados

Se han optimizado los metaparámetros utilizando un algoritmo genético que han sido programados en otro TFG. Un algoritmo genético consiste en un proceso de búsqueda de una solución basado en la simulación de la evolución genética.

Se han utilizado dos métodos de optimización: optimización multiobjetivo y optimización restringida y racha negativa 10%.

Por un lado, la optimización multiobjetivo es una función que busca maximizar la rentabilidad y minimizar la peor racha negativa de la rentabilidad total. Para ello se genera una función que pondera estos dos valores, siendo el peso otro metaparámetro que se va adaptando en cada operación. Los datos de rentabilidad y racha negativa para cada acción tras la optimización quedan explicados en las tablas 6.2 y 6.4 para K-means y GMM, respectivamente.

Por otro lado, la optimización restringida limita la racha negativa con respecto al beneficio total. La racha negativa total en este caso tiene que ser menor del 10%. En las tablas 6.3 y 6.5 se describen los valores obtenidos para K-means y GMM.

Se ha comprobado cada método (K-means y GMM) optimizando la siguiente lista de acciones.

Para optimizar se han utilizado los siguientes valores:

- Enagás
- MasterCard incorporated
- Kraft Heinz Company
- UnitedHealth Group

Para validar se han utilizado:

- Facebook
- Visa
- Abbvie Inc
- Abbot laboratorios

Los valores obtenidos en la optimización para cada metaparámetro aparecen en la tabla 6.1.

	Optimización multiobjetivo	Optimización restringida y racha negativa 10%
Ventana larga (días)	48	40
Ventana media (días)	32	36
Ventana corta (días)	20	35
Ventana futura (días)	32	8
Prefiltro	0.18	0.26
Umbral de compra (%)	-1	-1
Umbral de venta (%)	-1	-1

Tabla 6.1: Valores de los metaparámetros optimizados

Se va a utilizar un ratio para estudiar la rentabilidad obtenida por día invertido para cada una de las acciones. El cociente se define de la siguiente forma:

$$Ratio = \frac{Rentabilidad}{Días_inversión_activa} (\%)$$

Para calcular el ratio anualizado se multiplica el ratio anterior por 250 días.

En la tabla 6.2 se estudian los datos obtenidos tras la optimización y validación para la técnica K-means utilizando tanto la optimización multiobjetivo con racha negativa 0%.

	ENTRENAMIENTO				TEST			
VALORES	RENTABILIDAD (%)	RACHA NEGATIVA (%)	RATIO	RATIO ANUAL	RENTABILIDAD (%)	RACHA NEGATIVA (%)	RATIO	RATIO ANUAL
Enagás	60.6	-10.3	0.049	12.25	16.5	-14.5	0.0134	3.35
MasterCard Incorporated	322.8	0	0.265	66.25	238	0	0.196	49
The Kraft Heinz Company	20.2	0	0.0478	11.95	-53.3	-33.9	-0.205	-51.25
UnitedHealth Group Incorporated	130.5	-10.7	0.118	29.5	297.5	0	0.244	61
TOTAL	133.5			30	124.7			15.52
Facebook	359.3	0	0.468	117	52.1	0	0.064	16
Visa	164.3	0	0.141	35.25	201.4	0	0.165	41.25
Abbvie Inc	69.2	0	0.093	23.25	37.8	0	0.051	12.75
Abbot Laboratories	57.6	-0.75	0.0481	12.025	103.6	0	0.085	21.25
TOTAL	162.6			46.88	98.7			22.81

Tabla 6.2: Optimización multiobjetivo para K-means

En la figura 6.1 se representa la estrategia de inversión de la parte de entrenamiento de los valores utilizados en la optimización multiobjetivo de metaparámetros. En la tabla 6.2 se observa que las rentabilidades son altas y las rachas negativas son pequeñas comparando con la rentabilidad en cada caso.

La acción Enagás proporciona un 60.6% de rentabilidad con una racha negativa de -10.3% y siendo el número de operaciones mayor que en el resto de acciones. Al igual que Enagás, The Kraft Heinz Company genera un 20.2% de rentabilidad siendo su racha negativa de 0% y realizando un número considerable de operaciones. The Kraft Heinz Company produce un 130.5% de rentabilidad con una racha negativa de -10.7% y realizando operaciones suficientes.

Por el contrario, MasterCard Incorporated genera un 322.8% con una racha negativa del 0% debido a que realiza muy pocas operaciones lo que indica que se ha restringido mucho la racha negativa pero su rentabilidad anual es 66.25% siendo ésta la más alta entre todas las acciones utilizadas para optimizar los metaparámetros.

En la parte de test en el que se optimizan los metaparámetros y como aparece en la tabla 6.2 y en la figura 6.1, se observa un resultado de 124.7% de rentabilidad.

La racha negativa de las acciones MasterCard Incorporated y UnitedHealth Group Incorporated es 0% pero esto es debido a que el número de operaciones realizadas es pequeño con lo cual la racha negativa no varía mucho.

Teniendo en cuenta las acciones Enagás y The Kraft Heinz Company, las rachas negativas son muy altas comparadas con la rentabilidad que aportan, siendo estas 16.5% de rentabilidad con un -14.5% de racha negativa en el caso de Enagás y -53.3% con -33.9% de racha negativa en el caso de The Kraft Heinz Company.

En la parte de entrenamiento en la validación, el número de operaciones es considerable y las rentabilidades en relación con sus rachas negativas son buenas en general, con excepción Abbvie Inc que solo realiza 2 operaciones, proporcionando una rentabilidad final conjunta de 162.6%.

En la validación de la optimización multiobjetivo, el número de operaciones es muy pequeño en el test, esto justifica el 0% de racha negativa en la mayoría de los casos. Sin embargo, las acciones Facebook en el entrenamiento y Visa en entrenamiento y test producen una un ratio anual muy alto.

Utilizando el ratio definido anteriormente, se observa que el conjunto de acciones utilizadas para la optimización producen unos ratios anualizados de 30% y 15.52% en el entrenamiento y test respectivamente.

Si se estudian aquellas acciones utilizadas para la validación los ratios anualizados son 46.88% para el entrenamiento y 22.81% para el test.

Aunque en ambos casos difieran entre sí, estos valores anuales son muy buenos.

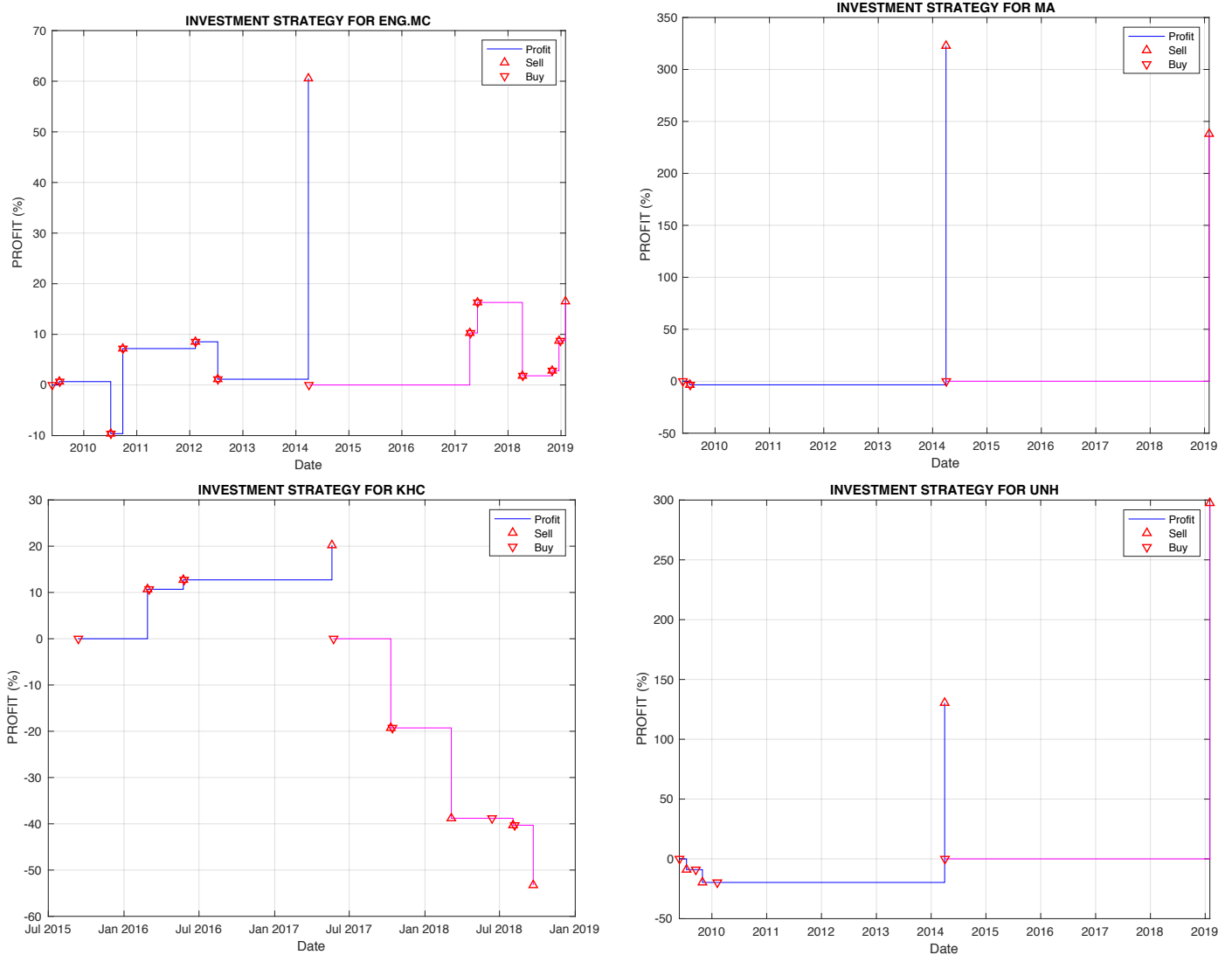


Fig. 6.1: Estrategia de inversión para las acciones utilizadas para la optimización multiobjetivo de metaparámetros para K-means

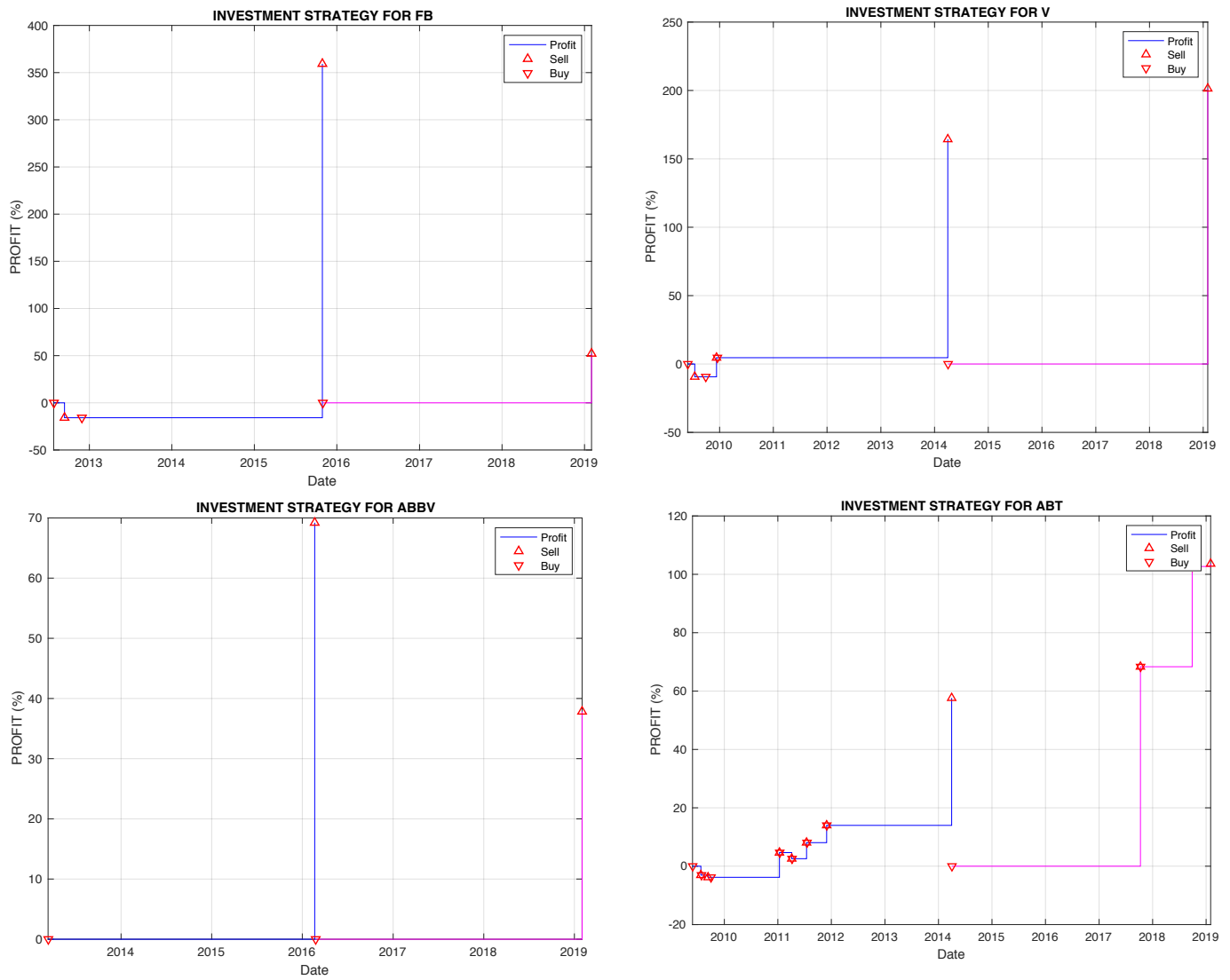


Fig. 6.2: Estrategia de inversión para las acciones utilizadas para la validación de metaparámetros en la optimización multiobjetivo para K-means

En la tabla 6.3 se estudian los datos obtenidos tras la optimización y validación para la técnica K-means utilizando la optimización restringida con racha negativa 10%.

	ENTRENAMIENTO				TEST			
VALORES	RENTABILIDAD (%)	RACHA NEGATIVA (%)	RATIO	RATIO ANUAL	RENTABILIDAD (%)	RACHA NEGATIVA (%)	RATIO	RATIO ANUAL
Enagás	34.3	-31.9	0.03	7.5	-8.7	-19.1	-0.0075	-1.88
Mastercard Incorporated	220.9	-15.2	0.185	46.25	292.2	0	0.237	59.25
The Kraft Heinz Company	10.1	-9.4	0.026	6.5	-61.2	-28.4	-0.217	-54.25
UnitedHealth Group Incorporated	113.8	-19.4	0.102	25.5	195.9	0	0.16	40
TOTAL	94.77			21.43	104.6			10.78
Facebook	358.6	-16.7	0.487	121.75	49.1	0	0.059	14.75
Visa	109.5	-16.4	0.095	23.75	261.7	0	0.212	53
Abbvie Inc	67.3	-3.6	0.09	22.5	34.8	-28.9	0.045	11.25
Abbot Laboratories	66.3	0	0.054	13.5	100.1	-18.5	0.083	20.75
TOTAL	150.44			45.37	111.5			24.94

Tabla 6.3: Optimización restringida y con racha negativa 10% para K-means

En la figura 6.3 se representan las estrategias de inversión de las acciones utilizadas para la optimización restringida de metaparámetros y con racha negativa 10%. En este caso el número de operaciones es mucho mayor comparado con la optimización multiobjetivo.

Tanto la acción Enagás como The Kraft Heinz Company no proporcionan una buena racha negativa en comparación con sus rentabilidades. Sin embargo, la acción UnitedHealth Group Incorporated proporciona un 113.8% de rentabilidad con un -19.4% de racha negativa en el entrenamiento y un 195.9% de rentabilidad con 0% de racha negativa en el test.

Por otra parte, Mastercard Incorporated genera un beneficio considerable en el entrenamiento, aunque la racha negativa en el test está muy restringida debido a que solo se realizan 2 operaciones.

En la figura 6.4 aparecen las estrategias de inversión de las acciones utilizadas para la validación.

El entrenamiento de Facebook genera una buena rentabilidad comparando con la racha negativa y teniendo en cuenta que se realizan un buen número de operaciones. En el caso de Visa esto se demuestra en la parte del test generando un 53% de ratio anual.

Teniendo en cuenta el ratio para ambos casos, con las acciones utilizadas para optimizar el ratio anualizado es de 21.43% en el entrenamiento y 10.78% en el test y, en el caso de las acciones utilizadas para la validación estos valores son 45.37% en el entrenamiento y 24.94% en el test.

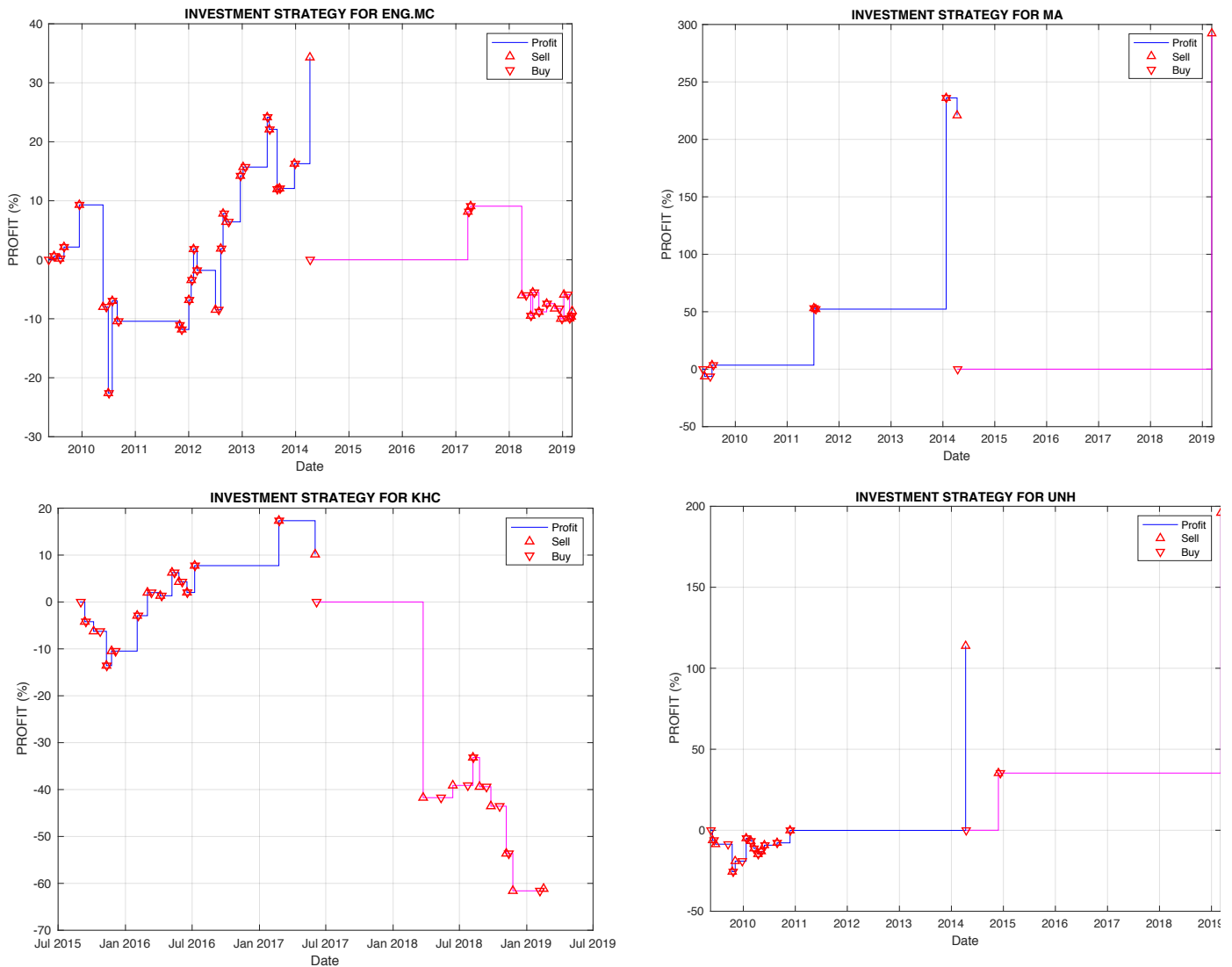


Fig. 6.3: Estrategia de inversión para las acciones utilizadas para la optimización restringida de metaparámetros para K-means

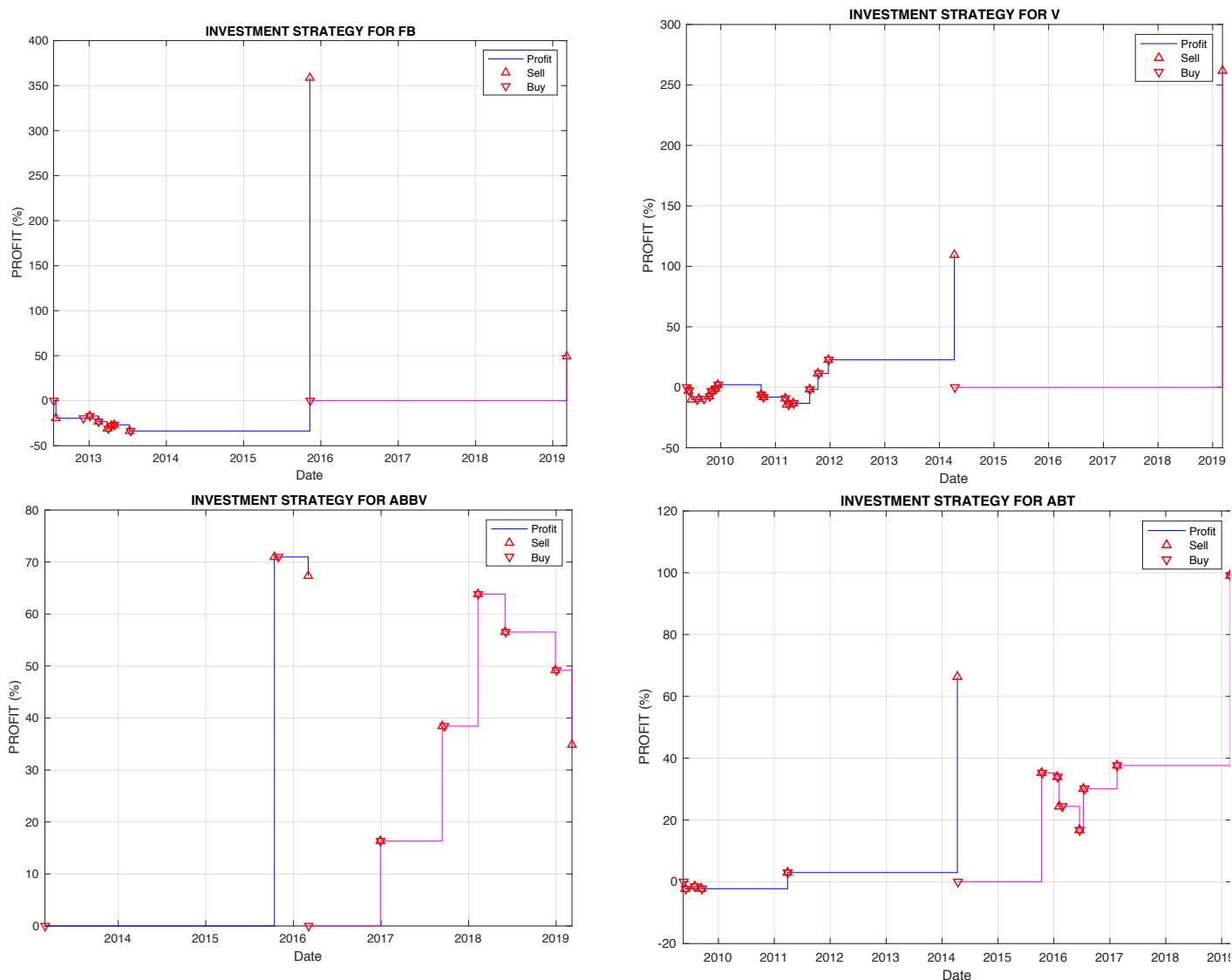


Fig. 6.4: Estrategia de inversión para las acciones utilizadas para la validación de metaparámetros en la optimización restringida para K-means

En la tabla 6.4 se describen los datos de la optimización multiobjetivo de los metaparámetros para el método GMM. En las figuras 6.5 y 6.6 quedan representadas las estrategias de inversión de dichas acciones con los metaparámetros optimizados, tanto para entrenamiento como para test.

	ENTRENAMIENTO		TEST	
VALORES	RENTABILIDAD (%)	RACHA NEGATIVA (%)	RENTABILIDAD (%)	RACHA NEGATIVA (%)
Enagás	53.94	-16.34	8.28	0
Mastercard Incorporated	236.87	0	238.01	0
The Kraft Heinz Company	21.6	-2.95	-37.2	-31.73
UnitedHealth Group Incorporated	120.03	0	297.48	0
TOTAL	108.11		126.64	
Facebook	301.86	0	52.15	0
Visa	157.58	0	201.36	0
Abbvie Inc	43.66	-19.54	28.12	-10.34
Abbot Laboratories	36.39	-17.02	129.22	0
TOTAL	134.87		102.71	

Tabla 6.4: Optimización multiobjetivo para GMM

En la tabla 6.4 aparecen marcadas en azul las únicas operaciones que aportan algún dato positivo debido a que el valor de su racha negativa es 0% y el número de operaciones no es demasiado pequeño o que el porcentaje de racha negativa en relación con la rentabilidad producida es bueno.

Aquellas cuya racha negativa es 0%, con excepción de las mencionadas previamente, no pueden estudiarse profundamente debido a que el número de operaciones realizadas es pequeño y esto produce que la racha negativa no varíe demasiado al estar muy restringida.

Las que tienen una racha negativa distinta de 0, con excepción de las mencionadas previamente, presentan un porcentaje de racha negativa muy grande en comparación con la rentabilidad que producen.

En conclusión, los datos conjuntos del entrenamiento tanto en las acciones utilizadas en optimización como validación son mucho más positivos que los obtenidos en el test.

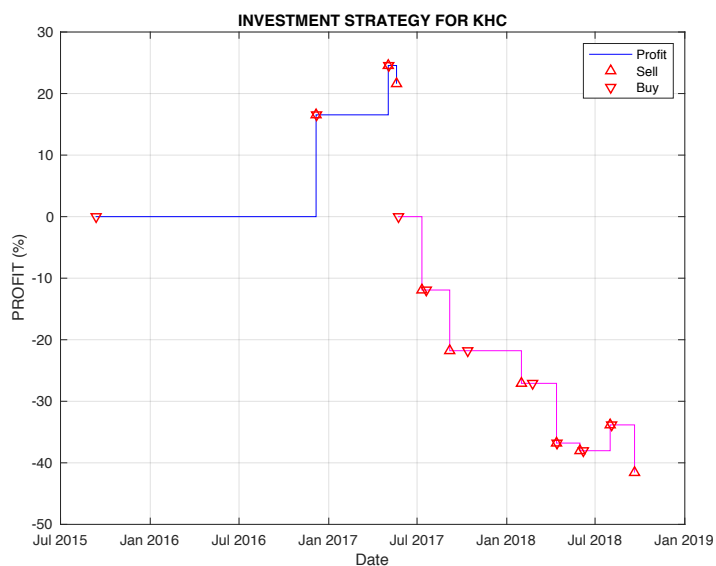
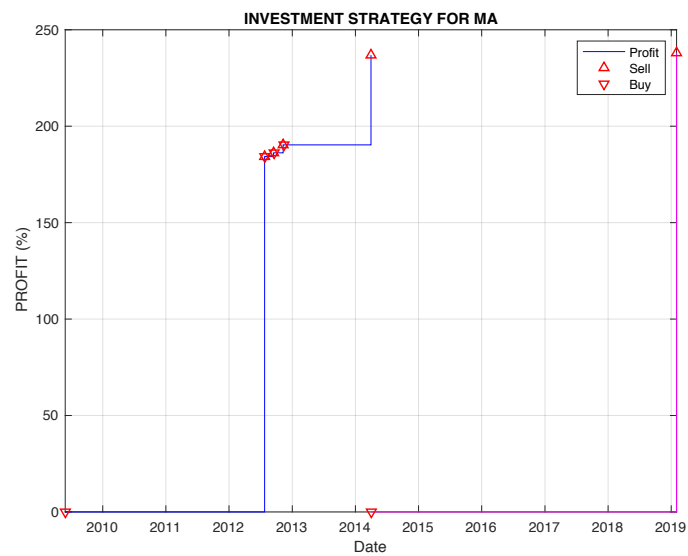
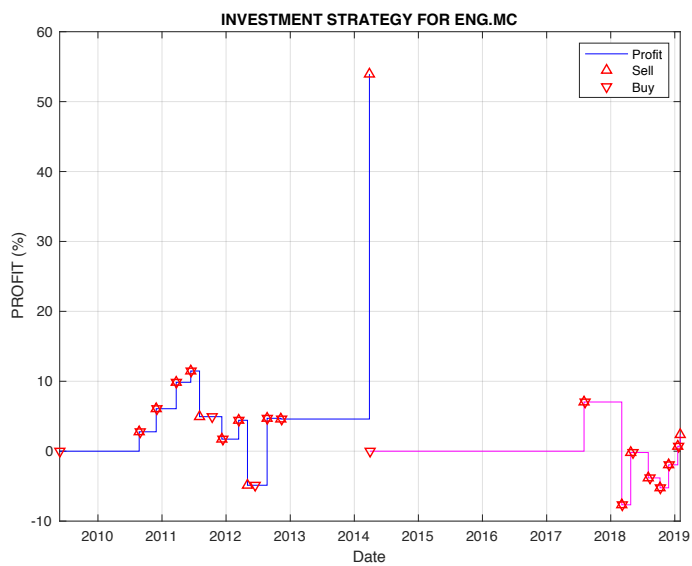


Fig. 6.5: Estrategia de inversión para las acciones utilizadas para la optimización multiobjetivo de metaparámetros para GMM

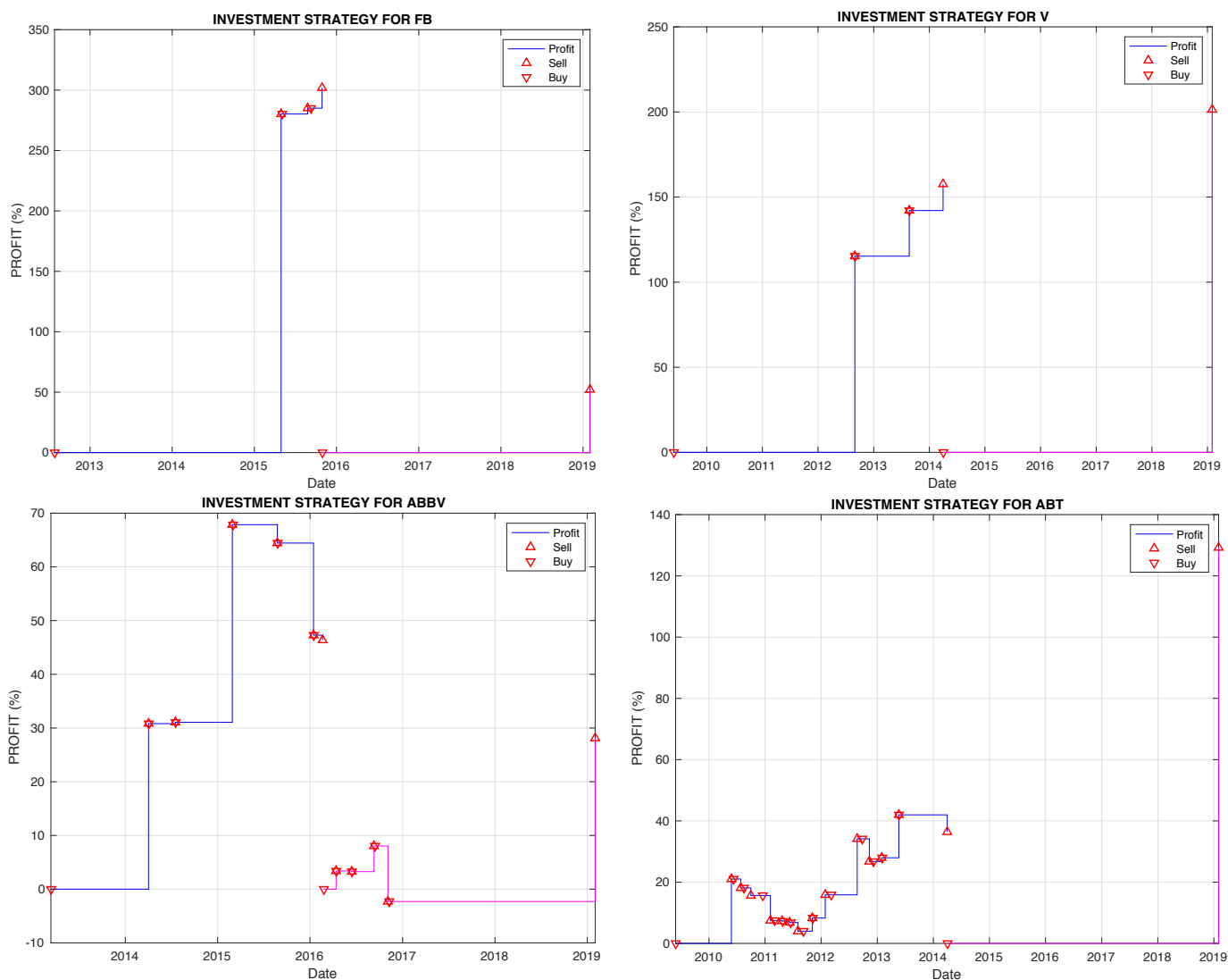


Fig. 6.6: Estrategia de inversión para las acciones utilizadas para la validación de metaparámetros en la optimización multiobjetivo para GMM

En la tabla 6.5 se describen los datos de la optimización restringida y con racha negativa 10% de los metaparámetros para el método GMM. En las figuras 6.7 y 6.8 quedan representadas las estrategias de inversión de dichas acciones con los metaparámetros optimizados, tanto para entrenamiento como para test.

	ENTRENAMIENTO		TEST	
VALORES	RENTABILIDAD (%)	RACHA NEGATIVA (%)	RENTABILIDAD (%)	RACHA NEGATIVA (%)
Enagás	-15.33	-52.73	4.25	-14.6
Mastercard Incorporated	92.43	-29.34	221.81	0
The Kraft Heinz Company	6.63	-11.7	-67.75	-57.69
UnitedHealth Group Incorporated	117.49	-15.52	200.57	0
TOTAL	50.31		89.72	
Facebook	119.76	0	46.44	0
Visa	59.68	-54.73	261.69	0
Abbvie Inc	33.66	-16.79	53.89	-12.96
Abbot Laboratories	10.71	-37.06	153.87	0
TOTAL	55.95		128.97	

Tabla 6.5: Optimización restringida y con racha negativa 10% para GMM

En el caso de la optimización restringida y racha negativa 10%, la mayoría de las rachas negativas de las acciones son muy altas en comparación con las rentabilidades que generan con excepción de UnitedHealth Group Incorporated en la optimización y Facebook en la validación.

Como conclusión general, este método aplicado al GMM produce peores resultados que el K-means.

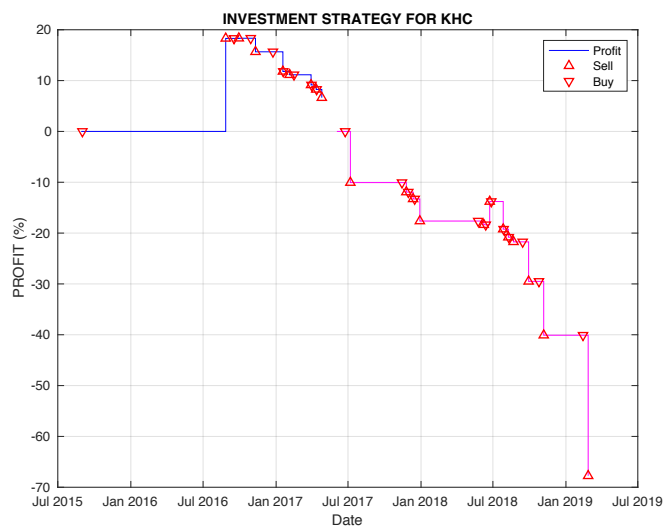
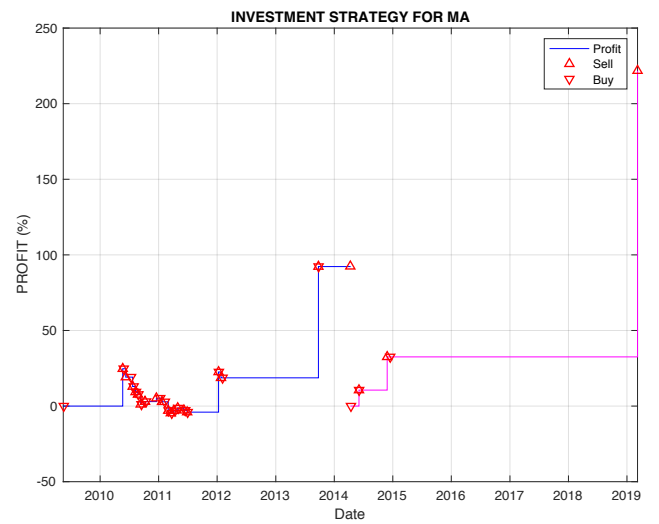
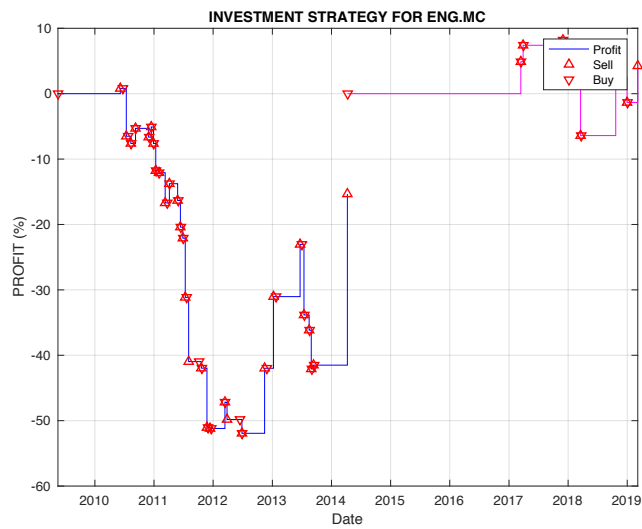


Fig. 6.7: Estrategia de inversión para las acciones utilizadas para la optimización restringida de metaparámetros para GMM

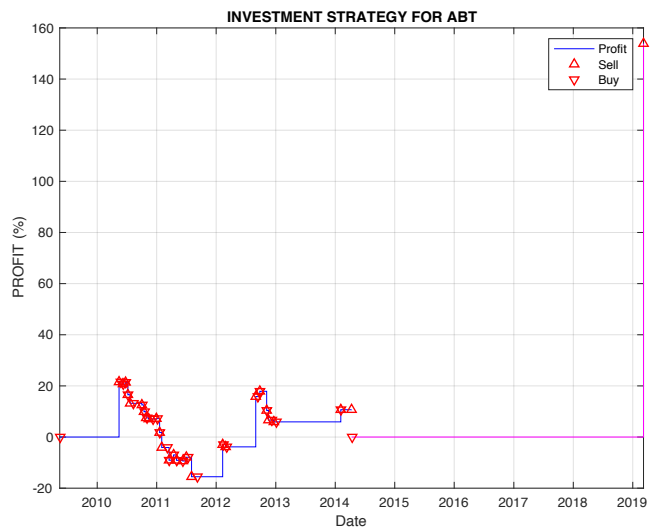
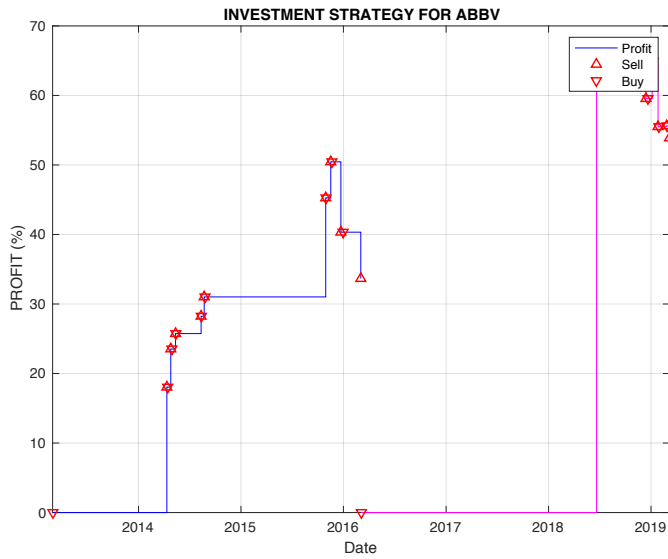
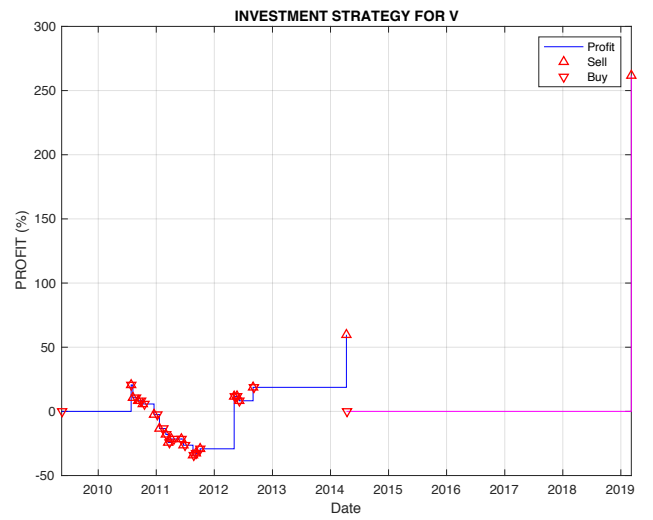
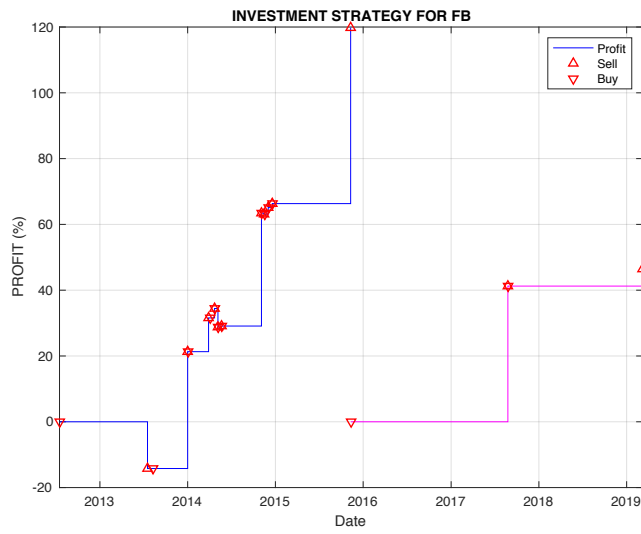


Fig. 6.8: Estrategia de inversión para las acciones utilizadas para la validación de metaparámetros en la optimización restringida para GMM

7. Conclusiones y futuros desarrollos

Tras haber programado los modelos de predicción para K-means y GMM y haber optimizado los metaparámetros para ambos métodos utilizando tanto optimización multiobjetivo como optimización restringida con racha negativa 10% se ha observado que ambos métodos funcionan muy bien para la técnica K-means mientras que para GMM presentan peores resultados.

La optimización multiobjetivo para el conjunto de acciones utilizadas en la optimización de metaparámetros para K-means (Enagás, MasterCard incorporated, Kraft Heinz Company y UnitedHealth Group) produce, como se muestra en la tabla 7.1, un 30% de ratio anualizado en el entrenamiento y un 15.52% de ratio anualizado en el test. Esta optimización aplicada al conjunto de acciones utilizadas para validación (Facebook, Visa, Abbvie Inc y Abbot Laboratories) genera unos ratios anualizados de 46.88% y 22.81% para entrenamiento y test respectivamente.

Optimización multiobjetivo	RATIO ANUALIZADO EN EL ENTRENAMIENTO	RATIO ANUALIZADO EN EL TEST
CONJUNTO DE VALORES USADOS PARA OPTIMIZACIÓN	30%	15.52%
CONJUNTO DE VALORES UTILIZADOS PARA VALIDACIÓN	46.88%	22.81%

Tabla 7.1: Rentabilidades anuales obtenidas en la optimización multiobjetivo para K-means

Por otro lado, como se muestra en la tabla 7.2, la optimización restringida con racha negativa 10% para el conjunto de acciones utilizadas en la optimización de metaparámetros para K-means genera un 21.43% de ratio anualizado en el entrenamiento y un 10.78% de ratio anualizado en el test. Esta optimización aplicada al conjunto de acciones utilizadas para validación genera unos ratios anualizados de 45.37% y 24.94% para entrenamiento y test respectivamente.

Optimización restringida y con racha negativa 10%	RATIO ANUALIZADO EN EL ENTRENAMIENTO	RATIO ANUALIZADO EN EL TEST
CONJUNTO DE VALORES USADOS PARA OPTIMIZACIÓN	21.43%	10.78%
CONJUNTO DE VALORES UTILIZADOS PARA VALIDACIÓN	45.37%	24.94%

Tabla 7.2: Rentabilidades anuales obtenidas en la optimización multiobjetivo para K-means

El conjunto de valores utilizados en la validación produce ratios anualizados muy parecidos con ambos métodos de optimización.

El conjunto de valores utilizados en optimización genera mejores resultados en la optimización multiobjetivo que en la optimización restringida pero ambos resultados son muy buenos.

Como futuros desarrollos se plantean los siguientes:

- Establecer un procedimiento para gestionar una cartera de valores decidiendo qué acciones se compran y cuáles se venden en cada momento.
- Estudios más minuciosos para la optimización de metaparámetros.

Bibliografía

- [1] Moreno, A., Armengol, E., Béjar Lluís Belanche, J., Cortés, U., Gavaldà, R., Gimeno, J.M., López, B. y Martín Miquel Sánchez, M.(1994). *Aprendizaje automático*. UPC: Barcelona. Recuperado en 29 de enero de 2019 de: <https://upcommons.upc.edu/bitstream/handle/2099.3/36157/9788483019962.pdf?sequence=1&isAllowed=y>
- [2] Aprende machine learning. (2017). *Aplicaciones del Machine Learning*. Recuperado en 29 de enero de 2019 de: <http://www.aprendemachinellearning.com/aplicaciones-del-machine-learning/>
- [3] Mathworks. (s.f.) *Machine learning*. Recuperado en 29 de enero de 2019 de: <https://www.mathworks.com/discovery/machine-learning.html>
- [4] Oswalso Lugo-Reyes, S., Maldonado-Colín, G y Murata, C. (2014). Inteligencia artificial para asistir el diagnóstico clínico en medicina. *Revista Alergia México*, 61(2), p. 110. Recuperado en 29 de enero de 2019 de: <https://nietoeditores.com.mx/nieto/Alergia/2014/abr-jun/Alergia2small.pdf#page=81>
- [5] Mercado, P. (2018) ¿Big Data y aprendizaje automático para invertir? *CincoDías*. Recuperado en 29 de enero de 2019 de: https://cincodias.elpais.com/cincodias/2018/06/15/mercados/1529067866_837590.html
- [6] Choudhry, R. y Garg, K. (2008). A Hybrid Machine Learning System for Stock Market Forecasting, *World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering* 2(3), p.689. Recuperado en 29 de enero de 2019 de: <http://waset.org/publications/8952> [Último acceso: 29 enero 2019]
- [7] Urraca Gesto, J.M.(s.f.) *Curso de análisis técnico*. Santander Central Hispano.
- [8] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15, 1929-1958. Recuperado en 29 de enero de 2019 de: <https://www.cs.toronto.edu/~hinton/absps/JMLRdropout.pdf>
- [9] Amatriain, X. (2016) *What are hyperparameters in machine learning?* Quora. Recuperado en 29 de enero de 2019 de: <https://www.quora.com/What-are-hyperparameters-in-machine-learning>

- [10] Kaushik, S. (2016). *An Introduction to Clustering and different methods of clustering*. Analytics Vidhya. Recuperado en 29 de enero de 2019 de: <https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/>
- [11] Nanda, S.R., Mahanty, B. y Tiwari, M.K. (2010), Clustering Indian stock market data for portfolio management. *Expert Systems with Applications*. 37(12), 8798-8798
Recuperado en 29 de enero de 2019 de: <https://www.sciencedirect.com/science/article/pii/S0957417410005300>
- [12] Zhang, Z. (2016) Selected Results from Clustering and Analyzing Stock Market Trade Data. *Kansas State University*. Recuperado en 29 de enero de 2019 de: <http://krex.k-state.edu/dspace/bitstream/handle/2097/39297/ZhihanZhang2018.pdf?sequence=3>
- [13] Mathworks. (s.f.) *Cluster Using Gaussian Mixture Models*. Recuperado en 29 de enero de 2019 de: <https://www.mathworks.com/help/stats/clustering-using-gaussian-mixture-models.html>
- [14] Marin, J.M (s.f.) Los mapas auto-organizados de Kohonen (SOM). Recuperado en 29 de enero de 2019 de: <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/tema5dm.pdf>
- [15] Enke, D., Grauer, M. y Mehdiyev, N. (2011), Stock Market Prediction with Multiple Regression, Fuzzy Type-2 Clustering and Neural Networks. *Procedia Computer Science*. 6, 201-206. Recuperado en 29 de enero de 2019 de: <https://www.sciencedirect.com/science/article/pii/S1877050911005035>
- [16] Sonnemans, J. (2006). Price clustering and natural resistance points in the Dutch stock market: A natural experiment. *European Economic Review*. 50(8), 1937-1950.
Recuperado en 29 de enero de 2019 de: <https://www.sciencedirect.com/science/article/pii/S0014292105001224>
- [17] Beyaz, E., Tekiner, F., Zeng, X. y Keane, J. (2018). Stock Price Forecasting Incorporating Market State. *IEEE*. Recuperado en 29 de enero de 2019 de: <https://ieeexplore.ieee.org/abstract/document/8623001/references#references>
- [18] Rostoker, C., Wagner, A. y Hoos, H. (2007). A Parallel Workflow for Real-time Correlation and Clustering of High-Frequency Stock Market Data. *IEEE*. Recuperado en 29 de enero de 2019 de: <https://ieeexplore.ieee.org/abstract/document/4227944>

[19] Martínez Olondo, J. (2017) Control predictivo para la gestión de una cartera de valores.

[20] Molero Senosian, A.L. (2018) Aplicación de técnicas de análisis de datos a la gestión optima de una cartera de valores