Applied Soft Computing, vol.180, pp.113300-1-113300-19, August, 2025.

Metric tools for sensitivity analysis with applications to neural networks

A. Muñoz San Roque; D. Alfaya Sánchez; J. Pizarroso Gonzalo; J. Portela González

Abstract-

As Machine Learning models are considered for autonomous decisions with significant social impact, the need to understand how these models work rises rapidly. Explainable Artificial Intelligence (XAI) aims to provide interpretations for predictions made by Machine Learning models, in order to make the model trustworthy and more transparent for the user. For example, selecting relevant input variables for the problem directly impacts the model’s ability to learn and make accurate predictions. One of the main XAI techniques to obtain input variable importance is the sensitivity analysis based on partial derivatives. However, existing literature of this method provides no justification of the aggregation metrics used to retrieved information from the partial derivatives. In this paper, a theoretical framework is proposed to study sensitivities of ML models using metric techniques. From this metric interpretation, a complete family of new quantitative metrics called α-curves is extracted. These α-curves provide information with greater depth on the importance of the input variables for a machine learning model than existing XAI methods in the literature. We demonstrate the effectiveness of the α-curves using synthetic and real datasets, comparing the results against other XAI methods for variable importance and validating the analysis results with the ground truth or literature information.

Index Terms- Sensitivity; Machine learning; Feature importance; Explainable A; Regressio; Feature engineering; Neural networks

Citation:
Alfaya, D.; Muñoz, A.; Pizarroso, J.; Portela, J. "Metric tools for sensitivity analysis with applications to neural networks", Applied Soft Computing, vol.180,

pp.113300-1-113300-19, August, 2025.