



Contribution ID: 140

Type: **not specified**

A Feature Selection Method Based on Shapley Values Robust to Concept Shift in Regression

Monday, 11 September 2023 12:00 (30 minutes)

Feature selection is one of the most relevant processes in any methodology for creating a statistical learning model. Generally, existing algorithms establish some criterion to select the most influential variables, discarding those that do not contribute any relevant information to the model. This methodology makes sense in a classical static situation where the joint distribution of the data does not vary over time. However, when dealing with real data, it is common to encounter the problem of the dataset shift and, specifically, changes in the relationships between variables (concept shift). In this case, the influence of a variable cannot be the only indicator of its quality as a regressor of the model, since the relationship learned in the training phase may not correspond to the current situation. Thus, we propose a new feature selection methodology for regression problems that takes this fact into account, using Shapley values to study the effect that each variable has on the predictions. Five examples are analysed: four correspond to typical situations where the method matches the state of the art and one example related to electricity price forecasting where a concept shift phenomenon has occurred in the Iberian market. In this case the proposed algorithm improves the results significantly.

Keywords

Concept shift, Feature selection, Regression

Classification

Mainly methodology

Primary author: SEBASTIÁN MARTÍNEZ-CAVA, Carlos (Fortia Energía - Universidad Politécnica de Madrid)**Co-author:** Prof. GONZÁLEZ GUILLÉN, Carlos (Universidad Politécnica de Madrid)**Presenter:** SEBASTIÁN MARTÍNEZ-CAVA, Carlos (Fortia Energía - Universidad Politécnica de Madrid)**Session Classification:** INVITED Spanish: Machine Learning in Business**Track Classification:** Data science