



Aprendizagem Automática Avançada (2º Ciclo de Informática)

Interpretability Exercises

Problem 1 *

For the following questions, please use the Breast Cancer Wisconsin (Diagnostic) as your dataset. Do not forget to split your data in train and test set. Consider a Random Forest Classifier.

- i) Explain the results obtained by this classifier by considering the Mean Decrease in Impurity (MDI).
- ii) Now consider the Permutation Feature Importance to explain the results obtained. In this case, compute and compare the results for the train and test sets.
- iii) Can you observe significant differences in the explanatory power in the two previous approaches in the training set? (see this link for more information).
- iv) Use SHAP and LIME to explain the results obtained by the model.
Note: You may find useful to take this tutorial on SHAP and LIME for tabular data (tip open in incognito window).

Problem 2 *

Take this tutorial on explaining image classification with LIME (tip open in incognito window).

- i) Use SHAP to obtain similar explanations to the ones you just obtained with LIME.
- ii) Which approach generated better explanations?