Universidade do Minho
Departamento de Informática
Mestrado [Integrado] em Engenharia Informática
Mestrado em Matemática e Computação

Dados e Aprendizagem Automática
1°/4° Ano, 1° Semestre
Ano letivo 2024/2025

Practical Exercise no. 6

Theme

Ensemble Learning

Exercise

Ensemble Learning is a ML technique that combines several models in order to obtain an optimal prediction model. There are several classes of Ensemble Learning, the most common being *BAGGing (Boostrap AGGregating), Boosting, Stacking* and *Max Voting (Majority Voting)*:

- **Bagging** involves adjusting decision trees to different samples from the same dataset, evaluating by calculating the average of these predictions.
- **Boosting** involves the sequential addition of ensemble members that correct the predictions made by previous models and produce a weighted average of the predictions.
- **Stacking** involves fitting different types of models to the same dataset, using another model to learn the best way to combine the predictions.
- **Max voting** involves fitting several models, each making a prediction and voting on each sample. Only the class with the highest votes is included in the final prediction class.

The aim of this practical statement is to carry out a series of tasks that will give you a better understanding of the structure and operation of ensemble learning models.

Tasks

The aim of this practical statement is for you to carry out a series of tasks that will give you a better understanding of the structure and operation of this ML technique. This practical exercise includes the following tasks:

T1. Follow and implement the steps presented in the lab notes in order to design the various ensemble learning models;

T2. Analyse and compare the results presented by each model in terms of their performance and predictions. List your conclusions;

T3. Try to vary the hyperparameters and combinations of the models and analyse how each of these variations influences the optimal prediction model;

T4. Follow and implement the steps presented in the theoretical notes in order to design some *ensemble learning* models within the *datasets* of the practical group work;

T5. Critically analyse the results obtained.