Machine Learning: Classification

1 Introduction

Python libraries required to complete the assignments:

- NumPy
- Matplotlib
- scikit-learn
- pandas

Introductory step:

- Load data from eportal to NumPy array (The best option is using Pandas).
- The first two columns represent the coordinates of the points (features) and the third one represents the label.
- Plot data.

2 Decision trees

Assignments

- Create a decision tree classifier.
- Fit the model with default values of hyperparameters.
- Perform prediction on training data.
- Calculate the accuracy.
- Play with hyperparameters such as *max_depth* and *min_samples_leaf*, and try to find the simplest possible model with the same or similar accuracy to the model with default hyperparameters.
- In the report explain how decision tree works and present results. Explain how hyperparameters *max_depth* and *min_samples_leaf* affect learning and what can cause the use of default hyperparameters. Describe the best solutions to prevent decision tree learning from overfitting. Discuss the pros and cons.

3 Support vector machines

Assignments

- Create several SVC classifiers with different kernels.
- Fit the models with different hyperparameters *C, kernel, degree, gamma*. (Try to achieve 100% accuracy).
- Perform prediction on training data.
- Calculate the accuracy.
- In the report, explain how SVM works, what a kernel trick is, and present the results. Explain the meaning of hyperparameters. Discuss the pros and cons.