



**Universidade do Minho**

Departamento de Informática

Mestrado [Integrado] em Engenharia Informática

Mestrado em Matemática e Computação

Dados e Aprendizagem Automática

1º/4º Ano, 1º Semestre

Ano letivo 2024/2025

Practical Exercise no. 10

**Theme** Support Vector Machines

**Exercise** Support Vector Machines (SVM) are supervised learning models applied to classification or regression problems, or outliers' detection, where the technique creates hyperplanes that separate the study cases into classes. The aim is to find a hyperplane that has the maximum margin, i.e. the maximum distance between the data points of both classes. This model allows linear and non-linear problems to be solved, producing models with significant accuracy and less computational effort.

The aim of this practical statement is to carry out a series of tasks that will give you a better understanding of Support Vector Machine models.

**Tasks** This practical exercise includes the following tasks:

**T1.** Follow and implement the steps presented in the practical notebooks to apply the Support Vector Classifier to the GenAudio dataset. Try to optimise the model by varying additional hyperparameters. You can use GridSearchCV to help choose the best set of hyperparameters. What is the best SVC configuration? How well the model performed, given an audio file as test case?

**T2.** Follow and implement the steps presented in the practical notebooks in order to design and optimize a SVM model in the context of the datasets of the practical group work.

**T4.** Given the results obtained in **T3**, what conclusions have you drawn? In which situations do the models get it right/fail? How can the proposed learning models be improved?

*Note:* The 'Private' attribute indicates the labelling of each university, showing whether the university is a private institute. For training purposes, this attribute should be *removed* from the dataset.

**T3.** Given the value of the 'Private' attribute, evaluate the clustering performance of each model by creating a confusion matrix (`sklearn.metrics.confusion_matrix(...)`) and a classification report (`sklearn.metrics.classification_report(...)`);