Ciências ULisboa | Informática

Aprendizagem Automática Avançada
(2º Ciclo de Informática)

Mixture Models Exercises

For the following problems, please use the Breast Cancer Winsconsin (Diagnostic) as your dataset (`https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html`).

Problem 1

Tip: scikit-learn has the `LogisticRegressionClassifier`, `DecisionTreeClassifier` and the `AdaBoostClassifier` class.

a) use a regression model to classify (logistic regression) and record the accuracy.

b) use AdaBoost with 7 logistic regression base learner experts and compare results.

c) compare with other numbers of experts (ex. 3, 10).

d) repeat a, b and c with decision tree base learner experts. Suggestion: try to use different number of experts, e.g. 15, 17, 20, 23, and so on.

Problem 2 *

Implement a Weighted Average Ensemble System in Python.

This ensemble system combines the output of several experts with a linear combination, whose weights are the accuracy scores of the experts on the dataset. This output should be rounded to the nearest integer.

Given a set of experts $\{f_1, f_2, \ldots, f_n\}$, with accuracy scores $\{a_1, a_2, \ldots, a_n\}$, and input X, the output of the system becomes:

$$\text{round}(\text{softmax}([a_1, a_2, \ldots, a_n]) \cdot [f_1(X), f_2(X), \ldots, f_n(X)]) \tag{1}$$

Use different types of classifiers as experts, with different hyperparameters, e.g. two Decision Trees with gini and entropy as criterion, two SVM's with polynomial and RBF kernel, and other classifiers you have previously learned.