Methods of machine learning

Exercise sheet III

May 14th, 2025

In this exercise class we apply boosting and bagging of decision trees to the classical *wines data set*. To this end, we use AdaBoost and random forests.

1. Task.

a) Load the data and split it into training and test sets (80-20 split).

b) Train a full decision tree classifier, check its training and test error and display the learned classifier.

c) Now apply the AdaBoost algorithm to this kind of learner (using `AdaBoostClassifier` from `sklearn.ensemble`) with $T = 10$ iterations. Check again training and test error and output the weights α_t, $t = 1, \dots, T$. What do you notice?

2. Task.

a) Now modify the decision tree learner to have at most $n = 2$ leaves. Again, train a corresponding decision tree classifier, check its training and test error and display the learned classifier.

b) Again apply AdaBoost $T = 10$ iterations. Check again training and test error and output the weights α_t, $t = 1, \dots, T$. What do you notice now?

c) Vary now the number of iterations $T \in \{1, \dots, 25\}$ and plot the resulting training and test error versus T. Which choice of T is best?

3. Task.

4. Task.