

Exercise Sheet 9

Exercise 1

Prove Condorcet's Jury Theorem using Chebyshev's inequality.

Exercise 2

Implement an `AdaBoost` class with a `fit` and `predict` method for binary classification problems. The `AdaBoost` class should accept any classifier from `sklearn` as a base learner that accepts `sample_weights` as an argument in its `fit` method.

Test your implementation against `sklearn`'s `AdaBoostClassifier` using `SAMME` as boosting algorithm on a dataset of your choice.

Exercise 3

Use `sklearn`'s digits dataset (`sklearn.datasets.load_digits`). Reassign the labels such that digits $0, \dots, 4$ are categorized as class -1 , and all other digits are categorized as class $+1$.

(a)

Randomly split the data into a training set (75%) and test set (25%). Apply AdaBoost using two different base learners:

- Decision Trees of depth 1
- Decision Trees of depth 5

For each base learner, determine a sufficiently large number of estimators such that AdaBoost nearly achieves a perfect training error. Plot the training and test error as functions of the number of estimators.

(b)

Apply AdaBoost with four different base learners using 10-fold cross-validation:

- Decision Trees of depth 1
- Decision Trees of depth 3
- Decision Trees of depth 5
- Decision Trees of depth 10

Plot the average training error of each AdaBoost variant as a function of the number of estimators.