University of Stuttgart | ANALYTIC COMPUTING

Foundations of Machine Learning - Exercise (SS 25)

Assignment 8: Random Forest and AdaBoost

Arvindh Arunbabu
arvindh.arun@ki.uni-stuttgart.de

Nadeen Fathallah
Nadeen.Fathallah@ki.uni-stuttgart.de

Cosimo Gregucci
cosimo.gregucci@ki.uni-stuttgart.de

Akram Sadat Hosseini
Akram.Hosseini@ki.uni-stuttgart.de

Farane Jalali
farane.jalali-farahani@ki.uni-stuttgart.de

Osama Mohammed
osama.mohammed@ki.uni-stuttgart.de

Jiaxin Pan
jiaxin.pan@ki.uni-stuttgart.de

Tim Schneider
tim.schneider@ki.uni-stuttgart.de

Jingcheng Wu
jingcheng.wu@ki.uni-stuttgart.de

Daniel Frank
daniel.frank@ki.uni-stuttgart.de

Submit your theoretical solution in ILIAS as a single PDF file.[1] Make sure to list the full names of all participants, matriculation number, study program, and B.Sc. or M.Sc on the first page. Optionally, you can *additionally* upload source files (e.g., PPTX files). Submit your programming task in ILIAS as a single Jupyter notebook. If you have any questions, feel free to ask them in the exercise forum in ILIAS.

Submission is open until Monday, 23rd of June, 12:00 noon.

[1]Your drawing software probably allows exporting as PDF. An alternative option is to use a PDF printer. If you create multiple PDF files, use a merging tool (like pdfarranger) to combine the PDFs into a single file.

Task 1: Random Forest

In the context of Random Forest, suppose that we have B decision trees: $T_1, T_2, \dots, T_B$. Show that the variance of the random forest predictor

$$\hat{f}_B(x) = \frac{1}{B}\sum_{i=1}^{B} T_i(x),$$

where each T_i is a decision tree, is given by Equation (1) as presented in the lecture:

$$\left(\rho + \frac{1-\rho}{B}\right)\sigma^2. \tag{1}$$

Here, σ^2 represents the variance of each individual T_i, and $\rho \equiv \rho(T_i, T_j)$ denotes the pairwise correlation between different decision trees (in particular, you may assume that this value is the same for all pairs). Recall that

$$\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Cov}(X_i, X_j) \tag{2}$$

$$\rho(X_i, X_j) = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{Var}(X_i)\mathrm{Var}(X_j)}} \tag{3}$$

Task 2: AdaBoost

Consider building an ensemble

$$f(x) = \text{sign}\left(\sum_{m=1}^{M} \alpha_m G_m(x)\right)$$

of decision stumps G_m using the AdaBoost algorithm. Figure 1 shows a set of labeled points in two-dimensional space. Initially, all data points are assigned equal weight of 1. A Jupyter helper script (`Adaboost_helper.ipynb`) is provided to simulate how AdaBoost updates the weights and selects decision stumps at each iteration.

> **Note:** In the Jupyter helper script, `polarity = 1` means that the decision stump classifies points **to the left or below the threshold as** $+1$, and the opposite side as -1. In the Jupyter helper script, Class $+1$ points are shown in red, and class -1 in blue.

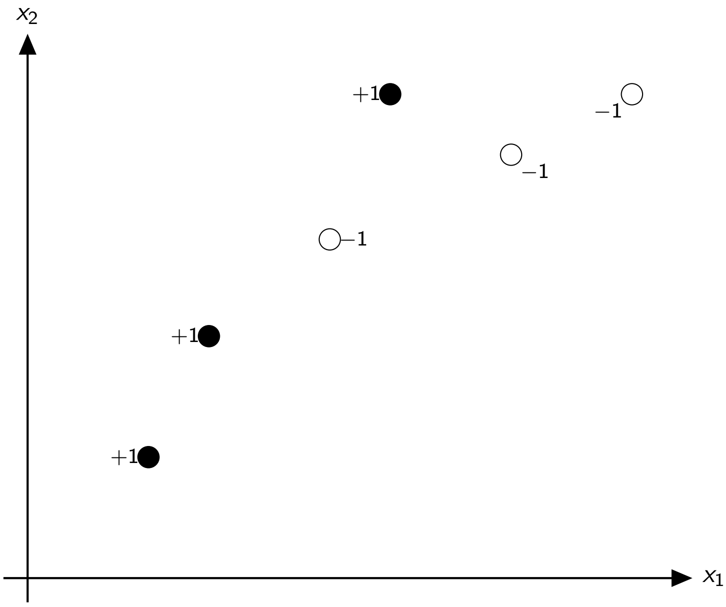

Figure 1 Labeled points for the AdaBoost algorithm.

1. **Task** Identify all the best initial decision stumps. These will serve as different starting points for running the AdaBoost algorithm.
2. **Task** For each of the best initial stumps identified, determine which data point(s) receive increased weights after the first iteration of AdaBoost. Provide the value of α_1 corresponding to each selected initial stump.
3. **Task** For each of the best initial stumps identified, determine which decision stump is selected in the second iteration of AdaBoost. What is the corresponding value of α_2? Compare α_1 and α_2, and interpret what this tells us about AdaBoost's learning process.
4. **Task** How does the final ensemble classifier differ for each of the three AdaBoost paths, depending on the initial stump selected? Describe how the choice of the first stump influences the resulting decision boundary and classification behavior.

Task 3: Random Forest Programming

Follow the instructions in the `08.ipynb` notebook.