Exercise Sheet 10

Exercise 1

Using the notation from the lecture, show that

$$W_e = \sum_{i=1}^{n} w_i^{(t+1)} 1_{h_t(x_i) \neq y_i} = \frac{1}{2}$$

for all $t \in \{1, \ldots, T-1\}$. Discuss this result.

Exercise 2

The goal of this exercise is to empirically validate some theoretical results of AdaBoost. For this, use the AdaBoost implementation from Assignment 9, Exercise 2. Let:

- $f^{(t)}(x) \in \{\pm 1\}$ be AdaBoost's decision function after t rounds.
- $\varepsilon_t^* = \varepsilon_t(h_t)$ be the error rate of the base learner at round t.
- $R_n\left(f^{(t)}\right)$ be the empirical risk of $f^{(t)}$.

(a) Run AdaBoost using decision stumps for $T = 10$ rounds. For each t print

- the sum of the weights of misclassified examples
- the sum of all weights $w_1^{(t+1)} + \cdots + w_n^{(t+1)}$
- the value $2\sqrt{\varepsilon_t^*(1-\varepsilon_t^*)}$.

(b) Run AdaBoost using decision stumps for T rounds. Plot the following metrics as a function of the number t of rounds:

- $\varepsilon_t^* = \varepsilon_t(h_t)$.
- $R_n\left(f^{(t)}\right)$.
- Bound of the empirical risk (see Theorem 9.6).

(c) Repeat exercise (b) using two different base learners:

1. A Decision Tree classifier of depth 10.
2. A random decision stump.

For each base learner, plot the same metrics as in part (b) as a function of the number t of rounds. Note that a random decision stump randomly selects a feature and a decision threshold.

Remark: Select an appropriate number T of rounds for each ensemble classifier to ensure a clear and informative visualization.

Exercise 3

Prove the following statement: AdaBoost's optimal decision function is of the form

$$f^*(x) = \arg \min_{f \in \mathcal{L}(\mathcal{B},T)} \mathbb{E}_{y|x}[\exp(-yf(x))] = \frac{1}{2} \ln \left(\frac{\mathbb{P}(y=+1|x)}{\mathbb{P}(y=-1|x)} \right)$$

or equivalently

$$\mathbb{P}(y=+1|x) = \frac{1}{1+\exp(-2f^*(x))}.$$

What does this statement remind you of?