Exercises for the course

Machine Learning 1

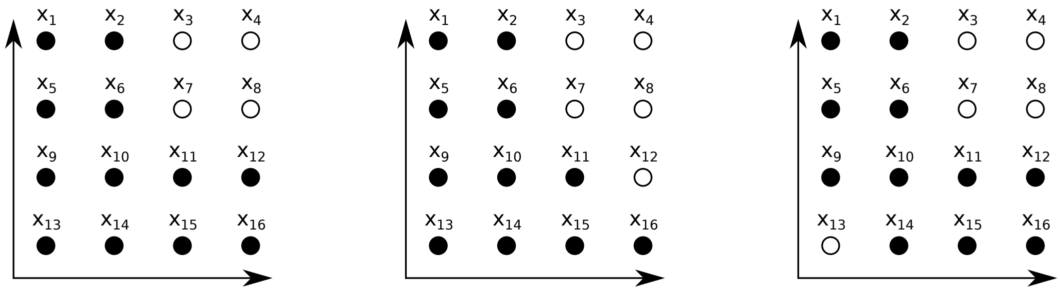
Winter semester 2018/19

Abteilung Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Exercise Sheet 13

Exercise 1: Boosted Classifiers (70 P)

We consider the following three two-dimensional binary classification datasets composed of 16 samples each:



Black circles denote the first class ($y_i = -1$) and white circles denote the second class ($y_i = 1$). We decide to use a boosted classifier with a perceptron as weak learner. The boosted classifier is given by

$$f(x) = \text{sign}\Big(\alpha_0 + \sum_{t=1}^{T} \alpha_t h_t(x)\Big)$$

where $\alpha_0, \dots, \alpha_T \in \mathbb{R}$, and where the function

$$h_t(x) = \text{sign}\big(w_t^\top x_i + b_t\big)$$

is the tth weak classifier, a simple perceptron with parameters w_t, b_t. Each weak classifier is trained to minimize the objective

$$\min_{w_t, b_t} \sum_{i=1}^{16} p_{t,i} \cdot \big[y_i - (w_t^\top x_i + b_t)\big]^2,$$

where the parameters $p_{t,1}, \dots, p_{t,16}$ have the role of weighting the data.

(a) *Build* at hand and for each dataset a possible boosted classifier, i.e. *draw* for each dataset the decision boundary of the weak classifiers $h_t(x)$ and of the final boosted classifier $f(x)$.

(b) *Write down* for each boosted classifier you have drawn, the coefficients $\alpha_0, \dots, \alpha_T$ and the weighting terms $p_{t,i}$ that have lead to each weak classifier.

Exercise 2: Boosted Regressors (30 P)

We consider the boosted regressor

$$f(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$$

where $\alpha_1, \dots, \alpha_T \in \mathbb{R}$, and where

$$h_t(x) = w_t^\top x$$

is the real-valued prediction produced by the tth weak regressor and $x \in \mathbb{R}^d$ with parameter w_t. Assuming a labeled dataset $(x_1, y_1), \dots, (x_N, y_N)$, the overall error of the boosted classifier is defined as:

$$\mathcal{E} = \sum_{i=1}^{N} \big[y_i - f(x_i)\big]^2$$

Like in the previous exercise, weak classifiers are learned by means of objectives of type:

$$\min_{w_t, b_t} \sum_{i=1}^{N} p_{t,i} \cdot \big[y_i - w_t^\top x_i\big]^2,$$

where the parameters $p_{t,1}, \dots, p_{t,N}$ have the role of weighting the data.

(a) *Show* that with $T = 1$ one can find a solution to this problem that cannot be improved by choosing $T > 1$.