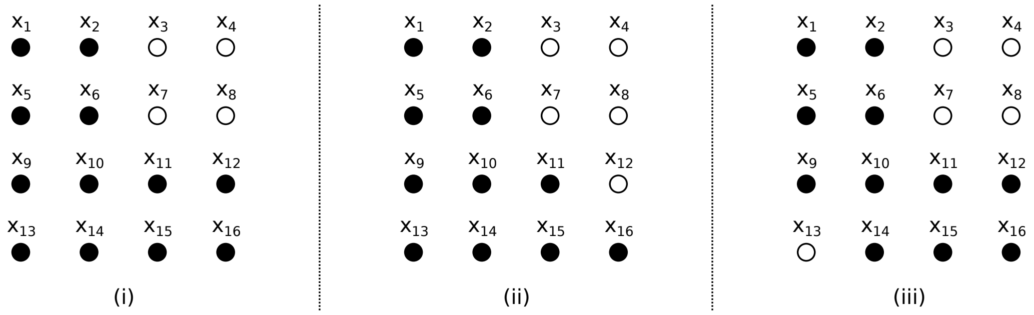
Exercises for the course
Machine Learning 1
Winter semester 2016/17

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: Weighting the Data (70 P)

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



Black circles denote the first class (-1) and white circles denote the second class ($+1$). We decide to use a boosted classifier with a linear soft-margin SVM as weak learner. The boosted classifier is given by the discriminant function

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

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

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

returns the classification result (-1 or $+1$) of the tth weak classifier. It is trained, under the weighting of the data $p_{t,1}, \dots, p_{t,16}$, to minimize the SVM objective

$$\min_{w_t, \xi_t, b_t} \frac{1}{2}\|w_t\|^2 + C \sum_{i=1}^{16} p_{t,i}\, \xi_{t,i}$$

under the constraints

$$\forall_{i=1}^{16} \;:\quad y_i(w_t^\top x_i + b_t) \geq 1 - \xi_{t,i}, \quad \xi_{t,i} \geq 0.$$

We also assume the parameter C to be large (e.g. $C = 100$).

Determine at hand and for each dataset a possible boosted classifier that classifies the data perfectly. *Draw* the decision boundary learned by each individual weak learner, and the final decision boundary. *Write down* the coefficients $\alpha_0, \dots, \alpha_T$, and the weighting terms $p_{t,i}$ for each weak learner $1 \leq t \leq T$ and data point $1 \leq i \leq 16$.

Exercise 2: Boosted Regressors (30 P)

We consider the boosted regressor

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

where $h_t(x) = w_t^\top x$ is the real-valued prediction produced by the tth weak regressor and $x \in \mathbb{R}^d$. Assuming a labeled dataset $\{(x_1, y_1), \dots, (x_n, y_n)\}$, the boosted regressor is trained to minimize the mean squared error

$$\sum_{i=1}^{n} \big(y_i - f(x_i)\big)^2, \tag{1}$$

where each weak regressor minimizes the following weighted objective function

$$\sum_{i=1}^{n} p_{t,i}\big(y_i - h_t(x_i)\big)^2.$$

Show that a single weak regressor can be made as accurate as the boosted regressor when using an appropriate weighting $\{p_1, \dots, p_n\}$. *Write down* one possible weighting for the single regressor that leads to the same accuracy as the optimal regressor in the sense of Equation (1).