Exam SCLASS 2011/12

(Dated: 27/01/2012)

Attention: Please sign and return also this question sheet! During the exam, books, lecture notes and calculators are allowed (but no devices connecting to internet).

General notation: We want to learn a classifier $\hat{C} : \mathcal{D} \to \mathcal{C}$ mapping d-dimensional feature vectors $\mathbf{x} = (x_1, x_2, ..., x_d) \in \mathcal{D}$ to a finite set of classes $\mathcal{C}$. The classifier shall be learned from the training set $\Pi_a = \{(\mathbf{x}^i, y^i) \mid i = 1, ..., N\} \subset \mathcal{D} \times \mathcal{C}$ of N correctly classified vectors $\mathbf{x}^i$.

Question 1: k-nearest-neighbor (k-NN) classification (6 points)

(a) Determine the training error (≡learning error) of a 1-NN classifier for an arbitrary training set!

(b) Assume a binary classification into classes $\mathcal{C} = \{\oplus, \ominus\}$, and $N_\oplus$ ($N_\ominus$) to be the number of positively (negatively) classified examples in Π_a. Determine the N-NN classifier $\hat{C}(x)$ for an arbitrary $x \in \mathcal{D}$.

(c) In Figure 1, the training set for a binary classification ($\mathcal{C} = \{\oplus, \ominus\}$) is denoted by red crosses ($\oplus$) and blue circles ($\ominus$). Use a 3-NN classifier to assign classes to the two orange data points (triangles) denoted as a and b in the figure. Show graphically in the figure, which data points have been taken into account to classify a and b.

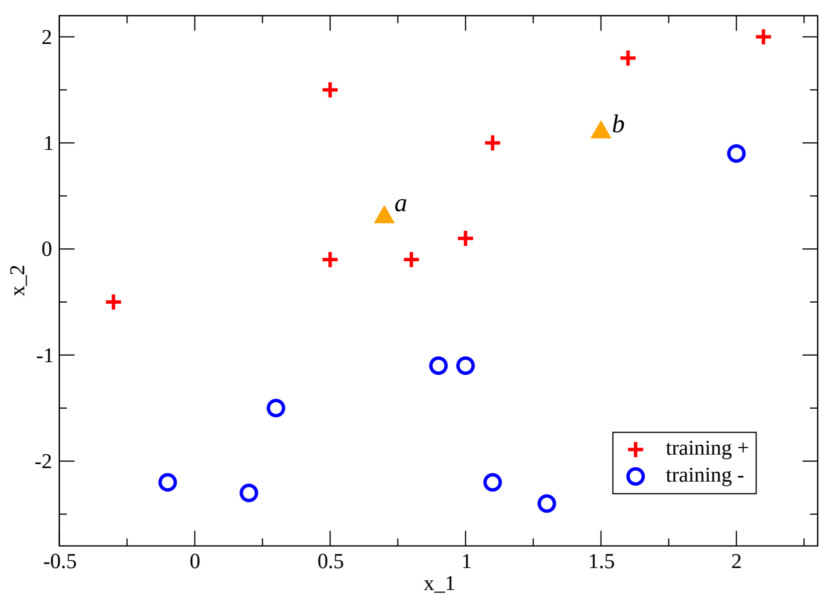

FIG. 1: 3-NN classification.

Question 2: Naive Bayesian classification (6 points)

(a) Many classifiers estimate statistical models $P(\mathbf{x}|y)$ for data points in each class, starting from Π_a. What is the specific assumption of Naive Bayesian classifiers, which makes this model learning feasible even for training set of moderate size N?

(b) Two classes of three-letter 'documents' are described by the following examples: $\mathbf{x}^1 = (A, A, B)$, $\mathbf{x}^2 = (B, B, B)$, $\mathbf{x}^3 = (A, B, C)$ for the first class, and $\mathbf{x}^4 = (C, C, D)$, $\mathbf{x}^5 = (D, D, A)$, $\mathbf{x}^6 = (D, B, C)$ for the second class. Construct, for each letter appearing in the two texts, and for each of the two classes, the probabilities $P(x|y)$, using a pseudocount 1.

(c) Determine the probabilities of the six training documents for each of the classes. Do they reflect correctly the class membership of each of the documents?

(d) Classify $\mathbf{x}^7 = (A, A, A)$ and $\mathbf{x}^8 = (A, B, D)$ according to the classifier constructed in (2.b+c). For which data point are you more confident, that the classification is actually correct?

Question 3: Binary linear classification (6 points)

(a) Is the training set Π_a in Figure 1 linearly separable? If yes, draw an example for a separating hyperplane into the figure.

(b) Given are the training set

$$\Pi_a = \{\ ((0,1), +1),\ ((0,-1), -1),\ ((1,1), +1),\ ((1,0), -1)\ \}$$

and an initial vector $\mathbf{w} = (-2, 1)$. Plot (schematically) the training set and the hyperplane defined by $\mathbf{w}' \cdot \mathbf{x} = 0$ ($w_0 = 0$). Determine the training error of $\hat{C}(\mathbf{x}) = \mathrm{sign}(\mathbf{w}' \cdot \mathbf{x})$.

(c) Use the perceptron algorithm to update $\mathbf{w}$, first using data point $(0, -1)$, second data point $(1, 1)$ (Remember: Each step of the perceptron algorithm uses one data point out of Π_a to update w). Use $\varepsilon = 0.5$. Draw the final updated hyperplane in the figure, and determine the updated learning error.

Question 4: Support vector machine (2 points)

(a) In Figure 2, two separating hyperplanes H_1 and H_2 are shown. Which one has to be preferred, if a support-vector machine shall be constructed?

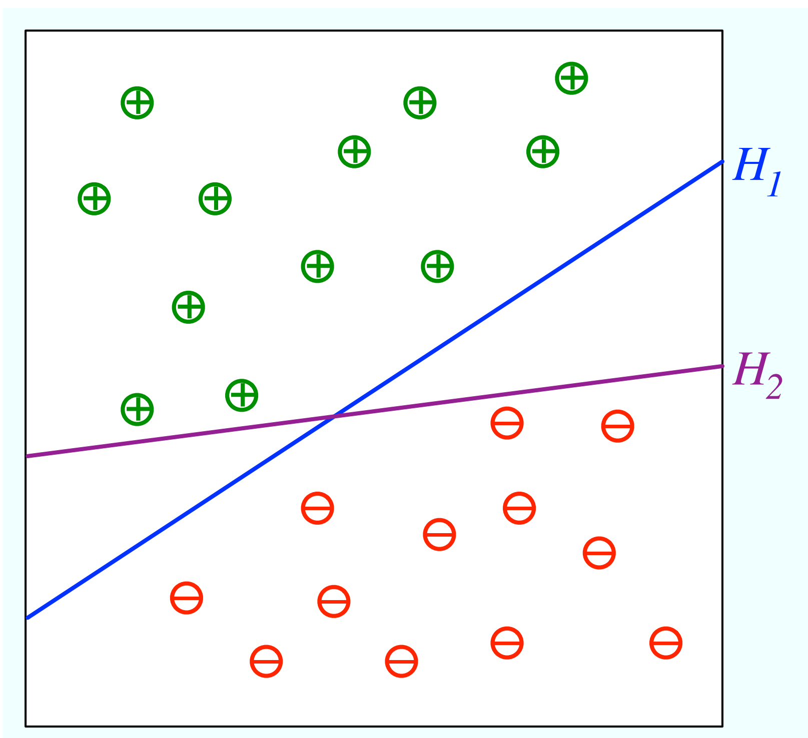

FIG. 2: Data set with two separating hyperplanes.