T-61.231 Principles of Pattern Recognition

Exercise 4: 14.10.2002

1. An attempt to estimate the error rate of the classifier is often made with the help of a test set (error percentage is calculated from the test vectors). Sometimes the test set is the same set with which the classifier was constructed. Show that this is not a good approach especially when the classifier is a 1-NN classifier.

2. Let us assume that two sets of points H_1 and H_2 in R^d are linearly separable. That means there is a vector $\underline{\omega}' \in R^d$ and a scalar $\omega_0 \in R^d$ so that

$$\underline{\omega}'^T \underline{x} - \omega_0 \begin{cases} > 0 & \forall \underline{x} \in H_1 \\ < 0 & \forall \underline{x} \in H_2 \end{cases}$$

   Let us augment the vectors in the following way:

$$\hat{\underline{x}} = \begin{pmatrix} \underline{x} \\ -1 \end{pmatrix}, \qquad \underline{\omega} = \begin{pmatrix} \underline{\omega}' \\ \omega_0 \end{pmatrix}.$$

   And let us finally change the sign of vectors $\hat{\underline{x}}$ of the set H_2.

   a) Show that $\underline{\omega}^T \hat{\underline{x}} > 0 \quad \forall \hat{\underline{x}}$.

   b) Show the situation graphically, when $d = 1$ and $H_1 = \{-3, -2, -1\}, H_2 = \{5, 6, 7\}$.

3. In the Perceptron Learning Rule, the updating $\underline{\omega}^{(n+1)} = \underline{\omega}^{(n)} + \alpha_n \hat{\underline{x}}_i$ is done, when $\hat{\underline{x}}_i$ has been classified incorrectly. That is $e_i = \underline{\omega}^{(n)T} \hat{\underline{x}}_i < 0$. How should α_n be chosen so that $\hat{\underline{x}}_i$ would surely be classified correctly after updating?

4. Let H_1 and H_2 be linearly separable, and the samples $x_1 = -2 \in H_1$, $x_2 = -1 \in H_2$ and $x_3 = 2 \in H_2$ are available for learning. Let the initial weight vector in the equation of question 3 $\underline{w}^0 = [1, 1]^T$ and the parameter $\alpha = 0.5$. Teach the percepton using these samples augmented as in question 2.