University of Stavanger

Stavanger, January 2, 2020

Theoretical exercise 4

ELE520 Machine Learning

Problem 1

We shall design a pattern recognition system to classify 2-dimensional feature vectors $\{\boldsymbol{x}\}$ to class ω_1 ('$\times$'), or class ω_2 ('$\circ$'). As a basis for this work, we have performed measurements so that 4 feature vectors are available as training vectors:

$$\mathcal{X}_1 = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix} \right\} \tag{1}$$

for class ω_1, and

$$\mathcal{X}_2 = \left\{ \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix} \right\} \tag{2}$$

for class ω_2.

a) Plot the training vectors in the x_1x_2–plane translated to the augmented feature space which will be used in the following. estimate the a priori probabilities.

b) We want to design a discriminant function on the form:

$$g(\boldsymbol{x}) = \boldsymbol{\theta}^T \boldsymbol{x} = \boldsymbol{y}. \tag{3}$$

where $\boldsymbol{x} = (x_1 \ x_2 \ 1)^T$. This is the same data set as in the example collection where straightforward linear regression was attempted and shown not to produce any result meaningful for discrimination.

Use the LS-method to determine $\boldsymbol{\theta}$. Determine and draw the decision boundary defined by $\boldsymbol{\theta}$.

c) What is the effect on the decision boundary of changing $\boldsymbol{y}_4$ to -0.5.

d) Use the LMS-method to determine $\boldsymbol{\theta}$. Choose the starting point $\boldsymbol{\theta}^{(1)} = (1\ 1\ 1)^t$. Set the threshold value to $\theta = 1$, and let $_{nm} = 0.5$. Remember that the learning rate is updated according to $_{nm}(i) =_{nm} /i$. Compute $\boldsymbol{\theta}^{(i)}$ for $i = 2, 3, 4, ...$ until convergence and sketch the decision boundary defined by $\boldsymbol{\theta}$. Compare the resulting decision boundary with the one you found in subtask a).

Trygve Eftestøl, Professor
Faculty of Science and Technology
Department of Electrical and Computer Engineering

Kjølv Egelands hus

University of Stavanger
N-4036 Stavanger.

Telephone: +47 51 83 10 00.
Telefax: +47 51 83 17 50.
E-mail: trygve.eftestol@uis.no
www.ux.his.no/~trygve-e

Problem 2

The problem of defining a generalised linear discriminant function to discriminante the vectors in a labelled two-class problem can be expressed

$$\boldsymbol{X}\boldsymbol{\theta} = \boldsymbol{y} \tag{1}$$

where $\boldsymbol{X}$ is the $N \times \hat{l}$ matrix[1] where the nth row is the vector $\boldsymbol{x}_n{}^T$ and $\boldsymbol{y}$ be the vector $\boldsymbol{y} = (y_1, \ldots, y_N)^T$.

a) Explain why this equation is not generally solvable.

b) Show that the vector $\boldsymbol{\theta}$ giving the best approximation in the sense of the minimum squared error $\|\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y}\|^2 = (\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y})^T(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y})$ between the right and left side in the equation is given by

$$\boldsymbol{\theta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{2}$$

c) Find an expression for the quadratic distance using the solution from the previous task.

d) Show that the problem is linearly separable if

$$\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T = \boldsymbol{I} \tag{3}$$

where $\boldsymbol{I}$ is the identity matrix.

Problem 3

Consider the data sets with samples $\boldsymbol{x}_i$.

$$\mathcal{X}_1 = \left\{ \begin{pmatrix} -3 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\} \tag{1}$$

for class ω_1, and

$$\mathcal{X}_2 = \left\{ \begin{pmatrix} -4 \\ -1 \end{pmatrix} \right\} \tag{2}$$

for class ω_2. We want to find the discriminant function $g(\boldsymbol{x}) = \boldsymbol{\theta}^T\boldsymbol{x}$ that solves the inequalities $\mathbf{a}^T\boldsymbol{y}_i > (<) 0$ for class $\omega 1(\omega_2)$.

The batch perceptron algorithm is given:

Algorithm 3. (Batch Perceptron)

[1] $\hat{l} = l + 1$

begin **initialize** $\boldsymbol{\theta}^{(\cdot)}$ $,_{nm}(\cdot)$, criterion θ, $i \leftarrow 0$
  do $i \leftarrow i + 1$
    $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} +_{nm} (i) \sum_{\boldsymbol{x} \in \mathcal{X}_i} y_n \boldsymbol{x}$
  until $|_{nm}(i) \sum_{\boldsymbol{x} \in \mathcal{X}_i} y_n \boldsymbol{x}| < \theta$
  return $\boldsymbol{\theta}$
end

a) Determine the labels, $y_n, n = 1, 2, \ldots, 3$ for each sample.

b) Plot the lines defined by $\boldsymbol{\theta}^T \boldsymbol{x}_i = 0$ for all the samples.

c) Indicate the positive and negative sides of the lines and identify the solution region.

d) Apply the algorithm letting the initial values be $\boldsymbol{\theta}^{(1)} = 0$ $,_{nm}(1) = 1$, criterion $\theta = 0$. Let $_{nm}(i) = 1$.

e) As above, but let let $_{nm}(i) = 1/i$.