

Stavanger, November 24, 2022

Theoretical exercise 3

ELE520 Machine Learning

Problem 1

We shall design a pattern recognition system to classify 2-dimensional feature vectors $\{\mathbf{x}\}$ 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:

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

for class ω_1 , and

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

for class ω_2 .

- 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.
- We want to design a discriminant function on the form:

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

where $\mathbf{x} = (x_1 \ x_2 \ 1)^T$. 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 $\boldsymbol{\theta}$.

- What is the effect on the decision boundary of changing y_4 to -0.5 .
- 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 $\mu = 0.5$. Remember that the learning rate is updated according to $\mu(i) = \mu/i$. Compute $\boldsymbol{\theta}^{(i)}$ for $i = 2, 3, 4, \dots$ until convergence and sketch the decision boundary defined by $\boldsymbol{\theta}$. Compare the resulting decision boundary with the one you found in subtask a). The resulting boundary will be a far shot from solving the problem.

You can do better with the the threshold value set to $\theta = 0.5$ and learning rate set to $\mu = 1$, but 16 iterations will be needed to reach convergence. If you have some time to spend and prefer this to watching TV, you might consider doing this.

Problem 2

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

$$\mathbf{X}\boldsymbol{\theta} = \mathbf{y} \quad (1)$$

where \mathbf{X} is the $N \times \hat{l}$ matrix¹ where the n th row is the vector \mathbf{x}_n^T and \mathbf{y} be the vector $\mathbf{y} = (y_1, \dots, y_N)^T$.

- Explain why this equation is not generally solvable.
- Show that the vector $\boldsymbol{\theta}$ giving the best approximation in the sense of the minimum squared error $\|\mathbf{X}\boldsymbol{\theta} - \mathbf{y}\|^2 = (\mathbf{X}\boldsymbol{\theta} - \mathbf{y})^T(\mathbf{X}\boldsymbol{\theta} - \mathbf{y})$ between the right and left side in the equation is given by

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

- Find an expression for the quadratic distance using the solution from the previous task.
- Show that the problem is linearly separable if

$$\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \mathbf{I} \quad (3)$$

where \mathbf{I} is the identity matrix.

Problem 3

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

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

for class ω_1 , and

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

¹ $\hat{l} = l + 1$

for class ω_2 . We want to find the discriminant function $g(\mathbf{x}) = \boldsymbol{\theta}^T \mathbf{x}$ that solves the inequalities $\mathbf{a}^T \mathbf{y}_i > (<) 0$ for class $\omega_1(\omega_2)$.

The batch perceptron algorithm is given:

Algorithm 3. (Batch Perceptron)

```

1 begin initialize  $\boldsymbol{\theta}^{(1)}, \mu(\cdot)$ , criterion  $\theta$ ,  $i \leftarrow 0$ 
2   do  $i \leftarrow i + 1$ 
3      $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} + \mu(i) \sum_{\mathbf{x} \in \mathcal{X}_i} y_n \mathbf{x}$ 
4   until  $|\mu(i) \sum_{\mathbf{x} \in \mathcal{X}_i} y_n \mathbf{x}| < \theta$ 
5   return  $\boldsymbol{\theta}$ 
6 end

```

Note that \mathcal{X}_i is the set of misclassified samples for iteration i .

- a) Determine the labels, $y_n, n = 1, 2, \dots, 3$ for each sample.
- b) Plot the lines defined by $\boldsymbol{\theta}^T \mathbf{x}_i = 0$ 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 $\boldsymbol{\theta}^{(1)} = 0, \mu(1) = 1$, criterion $\theta = 0$. Let $\mu(i) = 1$.
- e) As above, but let $\mu(i) = 1/i$.