University of Stavanger

Stavanger, December 17, 2024

Theoretical exercise 1

ELE520 Machine learning

Problems 4 and 5 are optional and are recommended for those who want to refresh some linear algebra and statistics. A PDF version of the student solution of the exercise shall be submitted on CANVAS.

Problem 1

A company producing sea food plans to start using an automated classification system to detect possible toxic content in blue mussels. The classification will be based on chemical analyses of one mussel from each container in a batch of incoming containers of newly caught blue mussels, and the concentration x of a specific chemical substance in the mussels will be used as feature for classification. If this concentration is above a specific threshold value, the whole container will be considered poisonous and rejected. The company's gain for each container is NOK 250,-, but the company has also issued a guarantee to its customers that if they buy a container that shows itself to be poisonous the company will compensate costs up to NOK 100000,-.

a) Let ω_1='toxic mussels' og ω_2='non toxic mussels'. Formulate the loss functions $\lambda(\alpha_i|\omega_j), i = 1, 2, j = 1, 2$, where α_i corresponds to the decision ω_i (risk will be the same as cost in this setting) when we do not include the possibility of rejecting classifications.

(You do not have to take into consideration that *correct classification* of non toxic mussels represents a gain which can be regarded as a "negative cost". Set the cost associated to this event to zero.!)

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

b) Dependent on the value of x we classify as ω_1 or ω_2. Express the conditional loss functions, $R(\alpha_i|x)), i = 1, 2$?

It is given that we know in advance that one of 25 containers contains toxic mussels, and that in toxic and non-toxic mussels the concentration x has a gaussian distribution $N(0.4, 0.0001)$ and $N(0.2, 0.0001)$ respectively.

c) Determine the decision boundary that minimises the overall loss (average cost) upon classification (R).

d) Determine the minimum average cost R upon classification in NOK.

Problem 2

We want to design a pattern recognition system that classifies 2-dimensional feature vectors $\{\boldsymbol{x}\}$ to class ω_1, or class ω_2.

It is known that the distribution between the two classes are 1/2 and 1/2 respectively for class ω_1 and class ω_2. Furthermore the feature vectors of the two classes are normally distributed around $\boldsymbol{\mu}_1 = (3\ 3)^t$ with covariance matrix

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix} \tag{1}$$

for class ω_1, og around $\boldsymbol{\mu}_2 = (3\ \ -2)^t$ with covariance matrix

$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \tag{2}$$

for class ω_2.

The eigenvalue and eigenvector matrices $\boldsymbol{\Lambda}_i$ and $\boldsymbol{\Phi}_i$, $i = 1, 2$ for the covariance matrices are

$$\boldsymbol{\Lambda}_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix} \tag{3}$$

and

$$\boldsymbol{\Phi}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{4}$$

for class ω_1, and

$$\boldsymbol{\Lambda}_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \tag{5}$$

and

$$\boldsymbol{\Phi}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{6}$$

for class ω_2 respectively.

a) Sketch the contour lines for the class specific probability density functions for the two classes in the same diagram.

b) Formulate Bayes decision rule for this problem. Determine the decision boundary and decision regions in the same diagram as in a).

Problem 3

Show that with

$$g_i(\boldsymbol{x}) = -\frac{\|\boldsymbol{x} - \boldsymbol{\mu}_i\|^2}{2\sigma^2} + \ln P(\omega_i) \tag{1}$$

as starting point where $i = 1, 2$, we can reach an expression defining the decision border between the two classes:

$$\boldsymbol{\theta}^t(\boldsymbol{x} - \boldsymbol{x}_0) = 0 \tag{2}$$

where

$$\boldsymbol{\theta} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j \tag{3}$$

and

$$\boldsymbol{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\sigma^2}{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2} \ln \frac{P(\omega_i)}{P(\omega_j)}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) \tag{4}$$

Problem 4

A similar problem can be found in the example problems set. The main difference lies in that while one-dimensional feature vectors are used in the example, you will have to solve the same kind of problem with two-dimensional feature vectors in this problem.

Given a random vector $\boldsymbol{x} = [x_1, x_2]^T$ characterised by the following two-dimensional probability density function:

$$p(\boldsymbol{x}) = \begin{cases} c & \text{if } a_1 < x_1 < b_1 \text{ og } a_2 < x_2 < b_2 \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The probability density function is also shown in figure 1.

a) Find c.

b) Find the expected value of $\boldsymbol{x}$ and make an illustration of $p(\boldsymbol{x})$ showing the value.

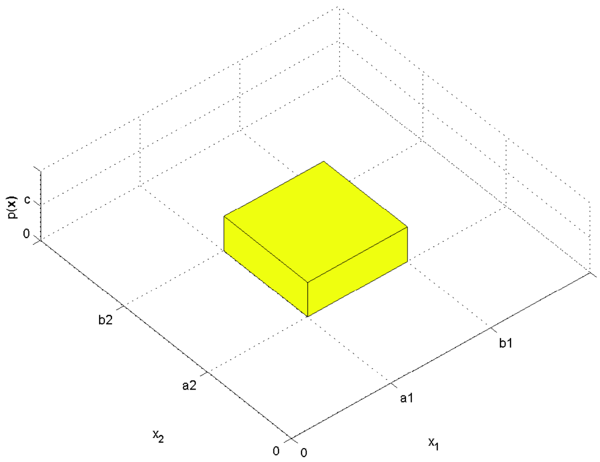

Figure 1: Uniform 2-dimensional probability density function.

c) Find the covariance matrix for $\boldsymbol{x}$. The following equation might be useful.

$$E[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T] = E[\boldsymbol{x}\boldsymbol{x}^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T \tag{2}$$

d) Explain the significance of a covariance matrix having identical elements on the diagonal.

e) Explain the significance of a covariance matrix being diagonal.

Problem 5

A similar problem can be found in the example problems set. This problem is given as an opportunity to freshen up your acquaintance with eigen vector analysis. This will later help recognise the significance of diagonalising the covariance matrix which is key to understanding how the data generated from a probability function is oriented in feature space where the principal directions are given by the eigen vectors and their magnitude by the eigen values.

A probability density function , $p(\boldsymbol{x})$, $\boldsymbol{x} = (x_1\ x_2)^t$, has a gaussian distribution around the expected value $\boldsymbol{\mu} = (1\ 1)^t$ with the covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix}. \tag{1}$$

a) Find the principal axes of the probability density function through decomposing $\boldsymbol{\Sigma}$ according to

$$\boldsymbol{\Sigma} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^t \tag{2}$$

der $\mathbf{\Lambda}$ og $\mathbf{\Phi}$ er eigenverdi- og eigenvektormatriser:

$$\begin{aligned} \mathbf{\Phi} &= \begin{pmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{pmatrix} \\ \mathbf{\Lambda} &= \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \end{aligned} \tag{3}$$

Arrange the eigen vectors according to the eigen values in descending order. Furthermore, the eigen vectors should have unity length ($\mathbf{e}_i^t \mathbf{e}_i = 1$). Principal axis number i might then be found by scaling the eigen vectors according to $\sqrt{\lambda_i}\mathbf{e}_i$.

b) Sketch the contour lines of $p(\boldsymbol{x})$ in the feature space spanned by $\boldsymbol{x}$. Indicate the principal axes and the expected value.