UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Exam in: TEK9020 – Pattern Recognition
Day of exam: 6.12.2021
Exam hours: 09:15 – 13:15
This examination paper consists of 4 pages.
Appendices: None
Permitted materials: No printed or handwritten material permitted. Approved, simple calculator permitted.

Make sure that your copy of this examination paper
is complete before answering.

Part 1

Introduction

a) Explain the concepts *class-conditional probability density function*, *prior probability* and *posterior probability*, and write down *Bayes rule* (Bayes formula) connecting these quantities.

b) Explain the principle of *minimum-error-rate* for choosing a class, and express this as a general decision rule.

c) Explain what a classifier is and describe the typical input data to the classifier.

d) Draw a figure showing the steps in a typical classification system, from raw data (measurements) to classification result.

Part 2

Decision theory

a) The multivariate normal distribution is given by

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right].$$

Explain the quantities involved. What are the parameters in this distribution?

b) In a two-dimensional problem with three classes, the class-conditional distributions are multivariate normal with common covariance matrix given by

$$\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}.$$

The mean (expectation) vectors for each class are

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ \boldsymbol{\mu}_2 = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \text{ and } \boldsymbol{\mu}_3 = \begin{bmatrix} -1 \\ 3 \end{bmatrix}.$$

Classify the feature vector

$$\boldsymbol{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

according to the principle og minimum-error-rate, when the prior probabilities of the classes are equal.

c) Draw a figure showing the mean vectors and the point $\boldsymbol{x}_0$ in feature space.

d) What is the shape of the decision boundaries between the classes in this case? Explain why. Sketch the decision boundaries in the figure.

Part 3

Parametric methods

a) Describe the *maximum-likelihood method* for estimating the parameter vector $\boldsymbol{\theta}$ in the assumed distribution function $p(\boldsymbol{x}|\boldsymbol{\theta})$ using supervised learning.

b) Write down the *likelihood function* $p(\mathscr{X}|\boldsymbol{\theta})$ and derive a system of equations for the estimate of $\boldsymbol{\theta}$ based on a training set $\mathscr{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ drawn from the given distribution function. Which assumption has to be made regarding these samples?

c) Derive the maximum-likelihood estimate of the parameter θ in the univariate distribution given by

$$p(x|\theta) = \frac{1}{6}\theta^4 x^3 \, \mathrm{e}^{-\theta x},$$

where $x \geq 0$ and $\theta > 0$. Assume a training set given by $\mathscr{X} = \{x_{1,\ldots,}x_n\}$.

Part 4

Linear discriminant functions

a) Write down a linear discriminant function $g(\boldsymbol{x})$ for a two class problem, and show how this can be rewritten in *augmented* form as the inner product of an augmented weight vector $\boldsymbol{a}$ and an augmented feature vector $\boldsymbol{y}$. Assume that the dimension of the original feature space is d. What is the dimension of the augmented feature space?

b) Describe how to train the weight vector $\boldsymbol{a}$ by using gradient descent, based on a training set of feature vectors. What is a *criterion function*? What is a *solution vector*?

c) Describe the *fixed-increment rule* for training the weight vector $\boldsymbol{a}$ using the training set $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$. What is required for this algorithm to terminate in a solution vector after a final number of iterations?

d) Assume a univariate training set consisting of the samples $1, 2, 6, 7$ where the first two samples belong to class ω_1 and the last two samples to class ω_2. Rewrite these samples in augmented form (remember the sign convention) and use the fixed-increment rule to find a solution vector. Use the initial weight vector $\boldsymbol{a}_0 = [0, 0]^t$.

e) Use this solution vector to find the decision boundary (threshold) between the classes i the original feature space.

Part 5

Nonparametric methods

a) Explain the difference between nonparametric and parametric methods, and write down an expression for the density estimate in a point $\boldsymbol{x}$ based on a training set of feature vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$.

b) Enter this estimate in Bayes formula to estimate the posterior probability for each class in a problem with c classes.

c) Explain how this estimate leads to the *k-nearest-neighbor rule* and express this decision rule in words.

d) Express the special case of the *nearest-neighbor rule* in words and write down an upper bound for the asymptotic error rate as a function of the optimal error rate.

Part 6

Unsupervised learning

a) What characterizes unsupervised learning (as opposed to supervised learning) and what is the meaning of a *mixture density*?

b) Write down the mixture density for a two-class problem in terms of the density functions and the prior probabilities for the individual classes.

c) Consider a *univariate* two-class problem. Use the maximum-likelihood method to show that the system of equations for the parameter vectors for the two classes can be expressed as

$$\sum_{k=1}^{n} P(\omega_i|x_k, \boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}_i} \ln p(x_k|\omega_i, \boldsymbol{\theta}_i) = 0, \quad i = 1, 2,$$

when the prior probabilities are assumed to be known. Here $P(\omega_i|x_k, \boldsymbol{\theta})$ is the posterior probability for class ω_i in the point x_k. The training set is given by $\mathscr{X} = \{x_1, ..., x_n\}$.

d) Assume further that the classes are normally distributed with equal prior probabilities and standard deviation equal to unity for both classes, while the mean values μ_1 and μ_2 are unknown. Derive a system of equations for the mean values and propose a solution method.