Department of Electronic Systems

Examination paper for TTT4275 Estimation, Detection and Classification

Academic contact during examination:: Magne Hallstein Johnsen
Phone: 93025534

Examination date: Date: Wednesday May 16th, 2018

Examination time (from - to): 09.00 - 13.00

Permitted examination support material: **C** – Specified, written and handwritten examination support materials are permitted. A specified, simple calculator is permitted

Other information:

- The examination consists of 3 problems where
  - problem 1 deals with estimation theory
  - problem 2 deals with detection
  - problem 3 deals with classification
- The point for each subproblem is given on top of the problem. The total number of points is 57.
- Grades will be announced 3 weeks after the examination date.

Language: English

Number of pages (front page excluded): 3

Number of pages enclosed:

Informasjon om trykking av eksamensoppgave			
Originalen er :			
1-sidig	☐	**2-sidig**	☐
sort/hvit	☐	**farger**	☐
skal ha flervalgsskjema			☐

Checked by:

Date Signature

Problem 1 Estimation (4+4+3+4+4=19)

In this problem we are interested in detectiong the amplitude A of a very weak astronomical signal signal, $A\cos(2\pi fn)$, where f is assumed known. The signal is embedded in background noise, $u[n]$, from other astronomical phenomena, so the actual observed signal is $A\cos(2\pi fn)+u[n]$.

In addition to the extrinsic noise, the equipment used to sample the signal also adds the noise $v[n]$ to the signal. Both noise sources are Gaussian, with $u[n] \sim \mathcal{N}(0, \sigma_u^2)$ and $v[n] \sim \mathcal{N}(0, \sigma_v^2)$, and are also independent.

1a) We are interested in estimating the parameter A from the signal

$$\begin{aligned} x[n] =& A\cos(2\pi fn) + u[n] + v[n] \\ =& A\cos(2\pi fn) + w[n], \ n = 0, \dots, N-1 \end{aligned}$$

where the frequency f is known, and $w[n] \sim \mathcal{N}(0, \sigma^2)$, $\sigma^2 = \sigma_u^2 + \sigma_v^2$. Find the probability density function for the observations $x[n]$.

1b) Show that the MVU estimator exists, and find it and the corresponding Cramer Rao lower bound.

We want to amplify our signal before sampling in an attempt to improve our estimate. We assume that we have an amplifier that is *almost* linear, that is, $y \approx Gx$, for some gain $G > 1$, where x is the input and y the amplified output.

1c) Write down the approximate model for $x[n]$ when the amplifier is used. Explain why the BLUE is a reasonable choice of estimator in this case.

1d) Find the BLUE and its variance based the approximate model for $x[n]$

1e) Assume that A is a random variable drawn from a known prior distribution. Define the Bayes mean square error estimate, and explain how you *in principle* would find this estimate using the information you have.

Problem 2 Detection (4+4+3+5+3 = 19)

After a long spring term at NTNU, you and your friends decide to go and hike in the Trondheim area. Unfortunately you realize too late that you forgot your water bottle at home. After a few hours of walking you find a river, which provides an opportunity to drink water. However, at the river bank is an information sign stating that *there is a suspicion that the water is polluted*. You are now faced with the following two hypotheses: Under hypothesis H_0 the water is drinkable and the remaining hike will be a blast, and; under hypothesis H_1 the water is polluted and will cause an acute stomach pain that will last for days. You are given the results of some water tests done along the river, which are summarized by a random variable x. If the water is fine, x is characterized by the following probability density function (PDF): $p_0(x) = e^{-x}u(x)$ with $u(x)$ being the unit step function. In case the water is polluted, x is characterized by the following PDF: $p_1(x) = xe^{-x}u(x)$. Attached to the sign is a test equipment which allows you to take a sample of the water.
You must now decide, based on the acquired sample value, whether or not to drink the water. Luckily you have studied TTT4275 and learned a little bit detection theory.

2a) Design a decision rule (including decision regions) such that the probability of detection P_D, i.e., the probability that you *decide the water is undrinkable* when the water is actually polluted, is maximized subject to the constraint that the probability of false alarm is $P_{FA} = 0.1$, i.e., you decide that the water is polluted when it is actually drinkable.

2b) Find the resulting P_D. Hint: $\int xe^{-x}dx = -e^{-x}x - e^{-x} + C$.

2c) Assuming that the probability that the water is polluted is $\pi_1 = 0.3$, find the test that will yield the minimum probability of error P_e.

2d) What is the probability of error in Problem 2c)?

2e) Which of the decision rules above is in your opinion more suitable for the problem at hand? Motivate your answer.

Problem 3 Classification : (3 + 3 + 4 + 5 + 4 = 19)

3a) Give the Bayes Decision Rule (BDR) for a C-class problem.

Use Bayes rule (BR) to rewrite BDR using class priors $P(\omega_i)$ and class densities $p(x/\omega_i)$.

The BDR classifier is optimal with respect to minimum error rate. Why is not possible to implement this BDR classifier?

3b) Given a 2-dimensional input room (observasion room). Sketch respectively i) a linear separable problem, ii) a nonlinear separable problem, iii) a nonseparable problem.

Give the decision rule for a discriminant classifier with $C \geq 2$ classes.

3c) Give the expression for a **linear** discriminant classifier.

Define the training cost function named sum of squared errors.

Explain why the training cost above requires use of sigmoids at the output.

3d) Explain the principle for a reference based classifier.

What is the difference between a NN-classifier and a KNN-classifier?

Give at least two different ways of finding references.

3e) Explain shortly the principle for clustering.