NORWEGIAN UNIVERSITY
OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF ELECTRONICS AND TELECOMMUNICATIONS

Contact during examination:
Name: Magne Hallstein Johnsen
Tel.: 93025534

EXAMINATION IN COURSE TTT4275
ESTIMATION, DETECTION AND CLASSIFICATION

Date: Wednesday May 16th, 2018
Time: 09.00 - 13.00

Permitted aids: –

INFORMATION

-

Problem 1 Estimation (e1+e2+e3+e4+5 = esum)

1a) Subtask E1

1b) Subtask E2

1c) Subtask E3

1d) Subtask E4

1e) Subtask E5

Problem 2 Detection (d1+d2+d3+d4+d5 = dsum)

2a) Subtask D1

2b) Subtask D2

2c) Subtask D3

2d) Subtask D4

2e) Subtask D5

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 class using 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?

Answer :
BDR : $x \in \omega_i \;\Leftrightarrow\; P(\omega_i/x) = max_k P(\omega_k/x)$

BDR + BR : $x \in \omega_i \;\Leftrightarrow\; p(x/\omega_i)P(\omega_i) = max_k p(x/\omega_)P(\omega_k)$

The priors and the densiities are not known. Thus we must either estimate both form and parameters (Plug-in-MAP) or choose another type of 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.

Answer :
See figure 4 in chapter 2.1 in course compendium titled "Classification".

Decision rule : $x \in \omega_i \;\Leftrightarrow\; g_i(x) = max_k \; g_k(x)$

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.

Answer :
Discriminant classifier : $g = Wx$ where g is a C-dimensional vector, x has dimension $D_x + 1$ (including offset) and W is a $Cx(D_x + 1)$ matrix

$SSE = \sum_n (t_n - g_n)^T (t_n - g_n)$ where g_n is the output due to x_n and t_n is the corresponding class label/target.

Since t_n is a binary target vector we have to squash the classifier outputs towards binary values. Thus we have $y_n = Wx_n \ \rightarrow g_n = sigmoid(y_n)$

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.

Answer :
We use a set of references with same (vector) dimension as the input. Each reference is labeled with class membership. The distances between the input and all references are calculated. The resulting distances are used for classification.

An NN classifier finds the closest (smallest distance) to the input and use the corresponding label for the input. A KNN classifier finds the K closest references. The majority class of the K references is chosen as input label.

References can be found in different ways : a) use all $N = \sum_i N_i$ available (and labeled) training observations, b) Draw randomly $L_i \ i = 1, \ldots, C$ references from each class. Use only a subset, i.e. $L_i << N_i$ of each class, c) C Perform clustering of the N_i training observations to find the L_i references.

3e) Explain shortly the principle for clustering.

Answer :
Clustering means to organize a set of N observations into a set L of clusters where $L << N$. In order to to this we have to decide upon a measure for the similarity/distance between an observation and a cluster (center). We start by choosing a number of clusters and intitial values for the cluster centers. We then "classify" all the observations and update the cluster (centers) based on the new labels. We do this classifying/updating procedure iteratively until no (or only small) improvements are made.

usually clustering is done in a hierarchical way, i.e. increasing the cluster numbers $L \rightarrow L + 1$ until a reasonable number of clusters is found.