Pattern Recognition
Exam on 2008-04-17

NO OPEN BOOK! GEEN OPEN BOEK! - It is not allowed to use the course book(s) or any other (printed, written or electronic) material during the exam.
Give sufficient explanations to demonstrate how you come to a given solution or answer!
The 'weight' of each problem is specified below by a number of points, e.g. (20 p).

1. (15 p) Maximum-likelihood estimation. Minimum error classification.
Consider a one-dimensional two-category classification problem with equal priors, $P(\omega_1) = P(\omega_2) = 0.5$, where the class conditional probability densities have the form

$$p(x|\omega_i) = 0 \text{ for } x < 0, \quad p(x|\omega_i) = \theta_i \exp(-\theta_i x) \text{ for } x \geq 0,$$

where θ_1 and θ_2 are positive but unknown parameters.

a) **(1 p)** Show that the distributions are normalized.
b) **(8 p)** The following data are collected: $D_1 = \{1, 5\}$ and $D_2 = \{3, 9\}$ for ω_1 and ω_2, respectively. Find the maximum-likelihood values of θ_1 and θ_2.
c) **(3 p)** Given your answer to part b), determine the decision boundary x^* for minimum classification error. Indicate which category is to the right and which to the left of x^*.
d) **(3 p)** What is the expected error of your classifier in part c)?

2. (10 p) *k*-means clustering.
Consider the application of the *k*-means clustering algorithm to the one-dimensional data set $D = \{0, 1, 5, 8, 14, 16\}$ for $k = 3$ clusters.

a) **(3p)** Start with the following three cluster means: $m_1(0) = 2$, $m_2(0) = 6$ and $m_3(0) = 9$. What are the values of the means at the next iteration?
b) **(5 p)** What are the final cluster means after convergence of the algorithm?
c) **(2 p)** For your final cluster means, to which cluster does the point $x = 3$ belong? To which cluster does $x = 11$ belong?

3. (10 p) Bayesian Estimation/Learning.
The task is to use Bayesian methods to estimate a one-dimensional probability density. The fundamental density function is a normalized triangle distribution $T(\mu,1)$ with center at μ and half-width equal 1, defined by

$$p(x|\mu) \sim T(\mu,1) = 1 - |x - \mu| \text{ for } |x - \mu| \leq 1,$$
$$= 0 \text{ otherwise}.$$

The prior information on the parameter μ is that it is equally likely to take any of the three discrete values -1, 0 or 1.

a) **(2 p)** Plot the 'estimated density' before any data are collected (which we denote by $D_0 = \{\}$). That is, plot $p(x|D_0)$. Here and below, be sure to label and mark your axes and ensure normalization of your final estimated density.

b) **(4 p)** The single point $x = 0.25$ was sampled, and thus $D_1 = \{0.25\}$. Plot the estimated density $p(x|D_1)$.
c) **(4 p)** Next the point $x = 0.75$ was sampled, and thus the data set is $D_2 = \{0.25, 0.75\}$. Plot the estimated density $p(x|D_2)$.

4. (5 p) Clustering. Dendrogram.
Construct a cluster dendrogram for the one-dimensional data D = {2, 3, 5, 10, 13}. As a distance between two clusters D_i and D_j use the maximum distance between a point from D_i and a point from D_j, for all possible pairs of such points.

5. (15 p) Binary decision trees.
This problem concerns the construction of a binary decision tree for three categories from the following set S of two-dimensional data:

ω_1: (1, 1), (2, 8), (3, 5); ω_2: (4, 7), (6, 2), (7, 6); ω_3: (5, 10), (7, 4), (7, 9) .

a) **(1 p)** What is the misclassifcation impurity at the root node, i.e., before any splitting?
b) **(4 p)** Let the query at the root node be "Is $x_1 > 3.5$?". What is the misclassifcation impurity drop achieved by this query?
c) **(8 p)** Continue to grow your tree fully. If two candidate splits are equally good, prefer the one based on x_1 (rather than x_2). Show the final tree and all queries.
d) **(2 p)** Use your tree to classify the point (7, 2).

6. (5 p) Learning vector quantization (LVQ). Describe the relevance LVQ algorithm.

7. (5 p) Receiver operating characteristics. What do you understand by a receiver operating characteristics (ROC)? To which class of problems does it apply? What is the common property of points that lie on the same ROC curve?

8. (5 p) Iris pattern recognition. Assume that you are given a set of 100 000 binary feature vectors, each of which is a binary code of the iris pattern of a person. The set contains 100 iris codes of each of 1000 persons. Describe how you would use this data to design an authentication system based on statistical decision theory.

9. (5 p) Eigenfaces. What is an eigenface and how is this concept used in face recognition?

10. (5p) Parzen windows.
Explain, using a simple example, the density estimation with Parzen windows. Under which conditions does this method give reliable results?