Pattern Recognition
Exam on 2007-04-16
Three hours, 100 points.

It is not allowed to use the course book during the exam. The use of print-outs of the lecture slides and the slides from student presentations as well as own notes is allowed.

1. Bayesian decision boundaries for normal distributions (30 points). Let us consider a two-category classification problem, with categories A and B with prior probabilities P_A and P_B. The class-conditional probability densities $p_{x|A}$ and $p_{x|B}$ are one-dimensional normal distributions:

$$p_{x|A} \sim N(\mu_A, \sigma^2_A), \qquad p_{x|B} \sim N(\mu_B, \sigma^2_B)$$

1a) Express analitically the position(s) of the optimal Bayesian decision boundary or boundaries in terms of P_A, μ_A, σ_A, P_B, μ_B, σ_B.

1b) Find the analytical conditions for having 0, 1, 2, or 3 decision boundaries. For each possible case, draw qualitative graphs of the posterior probability functions $P_A p_{x|A}$ and $P_B p_{x|B}$, which illustrate why the number of decision boundaries depends on the parameters P_A, μ_A, σ_A, P_B, μ_B, σ_B.

1c) Let us consider the sets of observations {-2, -1, 0, 1, 2} for category A and {3.2, 4.1, 5, 5.9, 6.8} for category B.

- 1c1) Compute *unbiased* maximum likelihood estimations of μ_A, σ_A, μ_B, σ_B.
- 1c2) Show that the condition $P_A = P_B$ gives rise to two decision boundaries and compute their positions x_1 and x_2.
- 1c3) Are both decision boundaries in 1c2 equally relevant? Justify your answer.

2. Leave-one-out (jackknife) estimation (15 points). Consider the following bag (multi-set) of numbers: {0, 3, 4, 5, 7, 10, 11, 11}. Using the leave-one-out (jackknife) procedure, estimate the median and the variance of the median.

3. Binary decision trees (20 points). Consider training a tree-based classifier with the following eight points of the form (x_1, x_2) from two categories

points from category w_1: (0, 4), (1, 2), (6, 0), (7, 3);
points from category w_2: (3, 5), (4, 7), (5, 6), (8, 1)

using misclassification impurity and queries of the form "Is $x_i < \theta$?" or "Is $x_i > \theta$?".

3a) What is the misclassification impurity at the root node (that is, before any splitting)?
3b) What is the optimal query at the root node? (Drawing the data will help you.)
3c) For the optimal query at the root node, what is the misclassification impurity at each of the immediate descendent nodes (that is, the two nodes at the next level)?
3d) Combine the misclassification impurities of these nodes to determine the impurity at this level. How much is the misclassification impurity reduced by your decision in part 3b?
3e) Continue splitting to create the full tree with 'pure' leaf nodes. Show you final tree, being sure to indicate the queries and the labels on the leaf nodes.

4. Minimum error classification. Missing features (20 points).
Consider a two-dimensional, three-category pattern classification problem, with equal priors $P(\omega_1) = P(\omega_2) = P(\omega_3) = 1/3$. We define the 'disk distribution' $D(\boldsymbol{\mu},r)$ to be uniform inside a circular disk centered on $\boldsymbol{\mu}$ and having radius r, and elsewhere 0. The class-conditional probabilities for the three categories are such disk distributions $D(\boldsymbol{\mu}_i,r_i)$, i = 1, 2, 3, with the following parameters:

ω_1: $\boldsymbol{\mu}_1 = (3, 2)$, $r_1 = 2$; ω_2: $\boldsymbol{\mu}_2 = (4, 1)$, $r_2 = 1$; ω_3: $\boldsymbol{\mu}_3 = (5, 4)$, $r_3 = 3$.

4a) **(4 points)** Classify the points (6, 2) and (3, 3) with minimum probability of error.
4b) **(16 points)** Classify the point (*, 0.5), where * denotes a missing feature.

5. K-nearest neighbour classification (15 points). Consider the following two sets of training patterns from two different classes:

Class 1: S_1 = {(7,31), (8,32), (10,32), (6,31), (5,32), (4,28), (5,30)}
Class 2: S_2 = {(10,31), (8,29), (9,33), (10,32), (14,33), (12,31), (11,30)}

Using Euclidian distance and 3-nearest-neighbour classification, classify the following test patterns: (8,31), (6,29), (10,30). If for a given test pattern more than three training patterns fall in its neighbourhood defined by the first three nearest neighbours from the training sets, use all training patterns falling in this neighbourhood to determine the class.