Pattern Recognition
Exam on 2008-06-23

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. Bayesian decision boundaries for normal distributions (20 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:

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

a) 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.

b) 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.

c) Let us consider the sets of observations {-3,-2, -1, 0, 1} for category A and {2.5, 3.3, 4, 4.7, 5.5} for category B.
   c1) Compute *unbiased* maximum likelihood estimations of μ_A, σ_A, μ_B, σ_B.
   c2) How many decision boundaries are there for $P_A = P_B$ and what are their positions?

2. (20 p) Minimum error classification. Missing features.
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: $\boldsymbol{\mu}_3 = (5, 4)$, $r_3 = 3$.

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

Hint: Draw the three disks and the points to be classified in a 2D feature space.

3. (10 p) *k*-means clustering.
Consider the application of the *k*-means clustering algorithm to the one-dimensional data set $D = \{0, 1, 5, 7, 8, 14, 16\}$ for $k = 3$ clusters.

a) **(3p)** Start with the following three cluster means: $m_1(0) = 2$, $m_2(0) = 4$ and $m_3(0) = 10$. 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 $x = 4$ belong? To which cluster does $x = 10$ belong?

4. (20 p) Binary decision trees. Consider the following multi-set S of two-feature patterns in a three-category problem. Each pattern is defined by a pair of features (f_1, f_2) where f_1 can take the values A or B and f_2 can take the values C or D. Each pattern is labeled by a category label w_1, w_2 or w_3. The labeled patterns in the multi-set S are:

S = {Patterns with label w_1: (A,C), (A,C), (A,C), (A,C);

Patterns with label w_2: (A,D), (B,C), (B,C), (A,D), (B,C), (A,D), (B,C);

Patterns with label w_3: (B,D), (B,D), (B,D), (B,D)}

a) Compute the misclassification impurity of S.

b) Split S in two multi-subsets L and R using the following rule and compute the impurity drop achieved by this split:

Q1: "Put a pattern in L if f_1 = A, otherwise put it in R."

c) Split S in two multi-subsets L and R using the following rule and compute the impurity drop achieved by this split.:

Q2: "Put a pattern in L if (f_1 = A AND f_2 = D) OR (f_1 =B AND f_2 = C) , else put it in R."

d) Which of the two rules Q1 and Q2 would you use for building a decision tree? Why?

e) Continue to grow your tree fully. Show the final tree and all queries.

5. (5p) Parzen windows.

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

6. (5 p) Present shortly the fuzzy k-means algorithm. What are the differences between k-means and fuzzy k-means?

7. (20 p) Hierarchical clustering. Consider the following set of binary patterns:

p_1=(1,1,1,1); p_2=(1,1,0,0); p_3=(1,0,1,0); p_4=(1,1,0,1);

p_5=(1,1,1,0); p_6=(1,0,0,0); p_7=(0,1,0,1); p_8=(1,0,1,1).

4a) Using the Tanimoto similarity, build a dendrogram and a Venn diagram for this set. The similarity between two sub-clusters is defined as the Tanimoto similarity of a pair of patterns, one pattern from one sub-cluster and the other from the other sub-cluster, for the pair for which a maximum similarity is reached. (A single pattern can also be considered as a sub-cluster.)

4b) What is the distance implied by this hierarchical clustering for the following pairs of patterns: (p_6,p_7), (p_2,p_4), (p_4,p_5). (The dissimilarity $\delta(p,q)$ between two patterns p and q is defined as $\delta(p,q)=1-s(p,q)$ where s(p,q) is their similarity.)

Reminder: Tanimoto similarity s(p,q) between two binary patterns p and q is defined as the ratio of the number of 1-bits that p and q have in common (p.q) and the number of bit positions in which either p or q has a 1-bit (p.p+q.q-p.q): $s(p,q) = p.q/(p.p+q.q-p.q)$.