IN4085
Pattern Recognition

Written examination
21-01-2011, 8.30-11.30

- There are four questions
- You have 30 minutes to answer the first question (Answer sheets 1-3), during which you cannot consult the book or any other material
- After you have handed in Answer sheets 1-3, you are allowed to consult any material you have brought to answer the following questions in the remaining 150 minutes
- Answer EACH QUESTION on a SEPARATE SHEET OF PAPER
- As much as possible, include the CALCULATIONS you made to get to an answer
- Do not forget to put your NAME and STUDENT NUMBER on top of every sheet
- Do not forget to hand this exam, including all Answer sheets

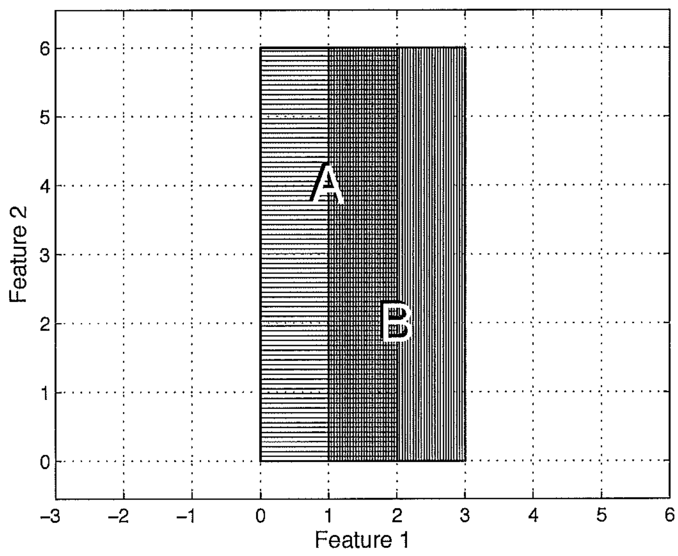
2 Classification *(10 points)*

a. Two equally probable classes A and B have both a 1-dimensional uniform distribution. Class A is distributed between 0 and 2. Class B is distributed between 1 and 6. Compute the Bayes error for this problem. *(1 points)*

b. Suppose the following training set is given for the above problem: 0.6, 0.8, 1.3 for class A and 1.7, 4.8, 5.8 for class B. Compute the classification errors for the nearest mean classifier, the nearest neighbor classifier and the linear support vector classifier. *(6 points)*

c. In a 2-dimensional problem (different from the one in case a.) for the two classes A and B the following training objects are given: [0,2], [0,6] for A and [4,2], [12,6] for B. As a classifier the majority voting combining classifier is used for three base classifiers: nearest mean, nearest neighbor and linear support vector classifier. Classify the following eight points using the combiner: [0,1],[0,3],[0,5],[0,7],[8,1],[8,3],[8,5],[8,7]. Hint: there is no need to formally compute the classifiers. A geometrically correct sketch is sufficient to solve the problem. It is not needed to hand in this sketch. Just the classifications of the 8 points will do. *(3 points)*

3 Feature Extraction and Selection *(10 points)*



Let class A be uniformly distributed on the rectangle defined by the corner coordinates (0,0), (0,6), (2,6), and (2,0). Class B is uniformly distributed as well, has the same shape as class A, but is shifted to the right. The coordinates defining its enclosing rectangle are (1,0), (1,6), (3,6), and (3,0). The priors of both classes are equal.

a. What is the Bayes error for this classification problem? *(1 points)*

b. Determine the total covariance matrix for the above 2-class problem. You may want to use that the within-class covariance [or within-class scatter] equals $\begin{pmatrix} 1/3 & 0 \\ 0 & 3 \end{pmatrix}$ and that the sum of the within-class scatter and the between-class scatter equals the total scatter. *(2 points)*

We now are going to reduce the dimensionality of this classification problem from the original two to a single one by means of principal component analysis [PCA].

c. Give the first, i.e., the most important, principle component. [Actual calculations may not be necessary as you may be able to see what the answer should be from the figure.] *(2 points)*

d. What is the Bayes error in the one-dimensional space defined by that first principal component? *(1 points)*

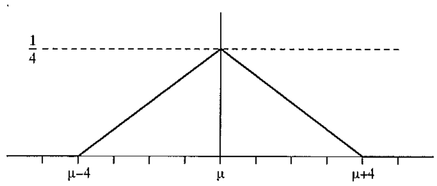
The foregoing was based on knowledge of the true, complete distribution, which compares to the situation in which we have an unlimited number of training examples. Obviously, in a more realistic pattern recognition setting, one has only a finite number of observations based on which to determine this primary principal component.

e. Describe and explain [qualitatively] how the expected Bayes error in the one-dimensional principal component space changes when the number of training samples for PCA decreases?

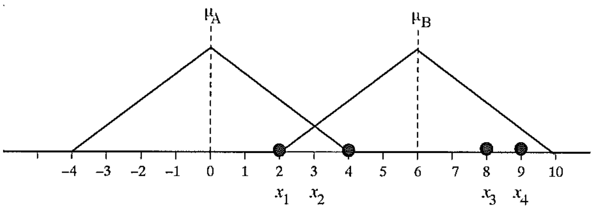
Does the expected performance become better or worse with less and less training examples. *(2 points)*

f. Substitute the Fisher mapping for PCA and reconsider e). What is now the overall expected qualitative behavior when the number of training samples decreases? Provide a brief explanation for this behavior? *(2 points)*

4 Clustering *(10 points)*



(a) Component PDF



(b) Initial configuration

Figure (a) above shows a univariate triangular probability density function with fixed width:

$$f(x;\mu) = \begin{cases} \frac{1}{4} - \frac{1}{16}|x-\mu| & \mu - 4 \leq x \leq \mu + 4 \\ 0 & \text{otherwise} \end{cases}$$

Note that this PDF has a single parameter, the mean μ.

A mixture of two such component PDFs, A and B, is fit to a dataset $\mathbf{X} = \{x_1, x_2, x_3, x_4\}$ using the EM-algorithm (Expectation-Maximization). The samples are shown in figure (b) (indicated by circles), together with the initial component PDF means μ_A and μ_B. The prior probabilities of both component PDFs are fixed at $\pi_A = \pi_B = 0.5$.

a. Determine the conditional probabilities $p(x_i|A)$ and $p(x_i|B)$ for $i = 1, 2, 3, 4$. *(2 points)*

b. Compute the loglikelihood of the data given initial configuration. *(1 points)*

c. Perform the E-step, i.e. calculate the posterior probabilities $w_{Ai} = p(A|x_i)$ and $w_{Bi} = p(B|x_i)$ for $i = 1, 2, 3, 4$. *(2 points)*

d. Perform the M-step, i.e. re-estimate the means μ_A and μ_B. *(2 points)*

e. What will be the final values of the means μ_A and μ_B when we re-iterate the E-step and M-steps many times? Please explain your answer. *(3 points)*

f. Calculate the Akaike Information Criterion (AIC) for the configuration you finally found. *(1 points)*