Pattern Recognition
Re-Examination on 2013-02-08

NO OPEN BOOK! GEEN OPEN BOEK! - It is not allowed to use the course book(s) or slides or any other (printed, written or electronic) material during the exam. You may only use a simple electronic calculator. 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. (1 p). You may give answers in English, Dutch or German language.

1) **Bayesian classification. (1 point)**
A medical test of a disease presents 2% false positives. The disease strikes 3 on 10000 of the population. People are tested at random, regardless of whether they are suspected of having the disease. A patient's test is positive. What is the probability of the patient having the disease?

2) **Bayesian Decision Theory. Normal distributions. (1 point)**
Given are the following two normal distributions:

$$p_1((x,y)\,|\,\omega_1) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2}\frac{(x-3)^2+(y-3)^2}{\sigma^2}\right)$$

$$p_2((x,y)\,|\,\omega_2) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2}\frac{(x-7)^2+(y-7)^2}{\sigma^2}\right)$$

We assume equal prior probabilities for the two concerned classes:

$$P(\omega_1) = P(\omega_2)$$

a) Find the decision boundary between the two classes in analytical form, i.e. as the simplest possible equation in x and y.
b) Draw this boundary and illustrate the two distributions by two circles centered on their respective means with radii of 1.
c) How can you estimate the classification error of this classifier? (If possible, write a mathematical expression.)

3) **MLE. (1 point)**
Consider the following set of feature vectors: S = {(1,1), (3,1), (1,3), (3,3), (2,2)}
Estimate the mean and covariance matrix of this data using unbiased maximum likelihood estimation (of the covariance matrix).

4) **Statistical decision theory, iris recognition. (1 point)**
Assume that an iris recognition system extracts iris code of 100 bits that are statistically independent. Assume that each individual bit can take the values 0 and 1 with equal probability (0.5) across the iris codes of different people. Estimate the probability that two iris codes of two different persons agree in more than 60% of the bits. Estimate the same probability for an iris code of 16 bits.

5) **k-nearest neighbors classification (1 point)**
Consider the following two sets of points in a 2D space:
A = {(2,3),(3,2),(3,3),(3,4),(3,5),(4,3),(5,3)}
B = {4,4),(4,6),(5,6),(6,6),(6,5),(6,7),(7,6)}

They come from two different classes ω_A and ω_B, respectively. Using Euclidean distance and the 3-nearest neighbors algorithm, classify the following test points: (6, 4), (6, 3), (2, 6), (4, 5). Hint: Computing distances may take you a lot of time. You can solve the problem faster by plotting the data and deciding visually which the nearest neighbors of a test point are.

6) **Hierarchical clustering (1 point).**

	O_2	O_3	O_4	O_5	O_6
O_1	4	5	6	5	1
O_2		5	6	2	3
O_3			1	4	6
O_4				5	6
O_5					4

The following upper triangular matrix describes the dissimilarities between six objects. Use the algorithm presented in the lectures to derive a dendrogram for these objects. Assume that the dissimilarity between two clusters of points is defined by the dissimilarity of their least dissimilar elements.

7) **K-means clustering. Vector quantization. (2 points)**

a) Present Lloyd's algorithm for k-means clustering.
b) Which function is minimized in the k-means problem?
c) Show that this algorithm converges.
d) Name problems encountered by Lloyd's algorithm for k-means clustering.
e) Describe the gap statistics method to determine the appropriate number of clusters in a data set.
f) Consider the application of the *k*-means clustering algorithm to the one-dimensional data set $D = \{-2, -1, 3, 6, 12, 14\}$ for $k = 3$ clusters. Start with the following three cluster means: $m_1(0) = 0$, $m_2(0) = 4$ and $m_3(0) = 7$. What are the values of the means at the next iteration? What are the final cluster means and clusters after convergence of the algorithm?
g) Sketch the vector quantization algorithm for clustering. What are the similarities and dissimilarities between k-means clustering and vector quantization clustering? Comment on their advantages and disadvantages.

8) **Relevance learning vector quantization (LVQ). (1 point)**

Assume that we deploy relevance LVQ with standard Euclidean distance for *N*-dimensional feature vectors that have to be assigned to 3 different classes.

a) Explain how the classification scheme is implemented by a relevance LVQ system with one prototype per class.
b) Explain relevance LVQ in terms of a few lines of pseudo-code. Consider a given set of training examples containing *N*-dimensional feature vectors and the corresponding class labels. Be precise and provide equations that define the update rules. If the update equations contain control parameters, explain their role.
c) Name and explain a method to evaluate the error of the LVQ classifier. How does this method differ from the method to be used in problem 2c above?

Math reminder:

Distribution of the normalized sum of n independent binary stochastic variables where p is the probability of 1 for a single binary variable: $N(p, p(1-p)/n)$.