# Pattern Recognition
## Examination on 2013-10-31

**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. (15 p). You may give answers in English, Dutch or German language.**

**1. Bayesian decision theory. Normal distributions. MLE. (40 points)**

The following two sets of feature vectors originate from two bivariate normal distributions:

$S_1 = \{(0, 0), (3, -1), (-1, 3), (2, 2), (1, 1)\}$

$S_2 = \{(3.5, 4.5), (3, 3), (5, 5), (4.5, 3.5), (4, 4)\}$

Problems:

a) Estimate the corresponding means and covariance matrices using maximum likelihood estimation. Use an unbiased estimator for the covariance matrices.

b) Find the analytical form of the optimal Bayesian decision boundary between the two classes, assuming the following relation between the prior probabilities $P(w_1) = 3P(w_2)$. (In this case, analytical form means a quadratic equation.) Draw a sketch of the boundary, together with the data sets and their estimated means.

c) Give a mathematical expression of the classification error of this classifier.

*Math reminders:*

a) The pdf of a multivariate normal distribution is defined as follows (*d* is the dimensionality):

$$p(\mathrm{x}) = \frac{1}{(2\pi)^{d/2} |\sum|^{1/2}} \exp\left[ -\frac{1}{2} (\mathrm{x} - \mu)^t \sum{}^{-1} (\mathrm{x} - \mu) \right]$$

b) Relation of a 2x2 matrix A, its determinant |A|, and its inverse $A^{-1}$:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad A^{-1} = \frac{1}{|A|} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$|A| = ad - bc$

c) Solutions of a quadratic equation $ax^2 + bx + c = 0$ (You may need this to determine the position of a point from the decision boundary):

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

2. **Learning vector quantization (LVQ). (15 points)**
   Assume that we deploy 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 an LVQ system with one prototype per class: $w^j$, j = 1,2,3. How does the *classification* process change when relevance LVQ is used?
   b) Explain the LVQ1 learning rule in terms of a few lines of pseudo-code. Consider a set of training examples (*N*-dimensional feature vectors with their corresponding class labels). Be precise and provide an equation that defines the update rule. If the update contains control parameters, explain their role.
   c) How does the *learning* process change when feature relevances are introduced? Provide an equation that defines the relevance update rule.
   d) What are the main differences between LVQ and k-nearest neighbor classification? What are their advantages and disadvantages?

3. **Statistical decision theory, iris recognition. (15 points)**
   Assume that you are given a set of 1 000 000 binary feature vectors, each of which is a binary code of the iris pattern of a person. The set contains 500 iris codes of each of 2000 persons.
   a) Describe how you would use this data to design an authentication system based on statistical decision theory.
   b) Describe the main components of hypothesis testing in statistical decision theory. How would you apply it to this problem?
   c) Assume that a complete iris code contains 2000 bits. Assume further that after image acquisition and analysis, 500 of the bits are missing. What do you need to change in your decision system and how much in order to achieve the same error of admitting an imposter as with a complete iris code that contains all bits.

4. **K-means clustering. Gap statistic. Vector quantization. (20 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 $D = \{-4, -3, 2.5, 5, 10, 12\}$ for $k = 3$ clusters. Start with the following three cluster means: $m_1(0) = -20$, $m_2(0) = -5$ and $m_3(0) = 8$. What are the values of the means at the next iteration? What are the final cluster means and clusters after convergence of the algorithm? Is there any problem with the initialization of the cluster prototypes? If the answer is yes, what solution would you propose?
   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.