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), (2,0), (0,2), (2,2), (1,1)}
   S_2= {(2,6), (3,3), (5,5), (6,2), (4,4)}

   Problems:

   a) Estimate the corresponding means and covariance matrices using unbiased
   maximum likelihood estimation (of the covariance matrices).

   b) Find the analytical form of the optimal decision boundary between the two
   classes, assuming the following relation between the prior probabilities P_2 =
   $2P_1$. (Analytical form means an equation.) Draw a sketch of the boundary,
   together with the data sets and the estimated means.

   c) How can you estimate the classification error of this classifier? (If possible,
   write a mathematical expression.)

   Reminders:

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

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

   Relation of a matrix A, its determinant |A|, and its inverse A^{-1}:

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

   $$|A| = ad - bc$$

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.

b) Explain LVQ1 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 an equation that defines the actual update rule. If the update contains control parameters, explain their role.

c) Explain a method to evaluate the error of the LVQ classifier. How does this differ from the method to be used in problem 1c above?

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, half of the bits are missing (due to various reasons). What do you need to change in your 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. 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 $D = \{-3, -2, 2, 5, 11, 13\}$ for $k = 3$ clusters. Start with the following three cluster means: $m_1(0) = -1$, $m_2(0) = 3$ and $m_3(0) = 6$. 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.