Pattern Recognition
Examination on 2011-11-01

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. 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).

1. **(20 p) Support Vector Machines (SVM).** Make a drawing of a two-dimensional (2D) classification problem in which you show two (training) sets of data points – black points that belong to a class ω_1 and white points that belong to a class ω_{-1}. Draw the two sets of points such that they are linearly separable.
   a) What different types of linear classifiers can you use for this classification problem? What are their advantages and disadvantages? Which are your arguments in support of or against using one type of classifier versus another type.
   b) Draw a linear decision boundary between the two sets of points and explain the concept of a margin in the context of SVM.
   c) Let the decision boundary be specified by the equation $\mathbf{w}^T.\mathbf{x}-b=0$ where $\mathbf{w}$ is a 2D orientation vector, $\mathbf{x}$ is a 2D vector that specifies the coordinates of a point in the feature space, and b is a constant. Give an expression for the distance of a data point $\mathbf{x}_i$ from the training sets to the decision boundary. Give an expression for the minimal distance from such a data point to the decision boundary, across all data points from the training data sets. Subsequently, give an expression for the margin. Next, give an expression for the maximum possible margin.
   d) The margin function considered above needs to be maximized but this process is subject to certain restrictions. Which are these restrictions?
   e) Consider now the following two 1D data sets: $S_1 = \{-11, -10, -9, -3, -2, -1, 0, 1\}$, $S_{-1} = \{-7, -6, -5, 3, 4, 5, 6, 7\}$. How would you map these 1D data points to 2D data points so that the two new (2D) data sets become linearly separable?

2. **(15 p) Hierarchical clustering**. The following upper triangular matrix describes the dissimilarities between six objects.

	O_1	O_2	O_3	O_4	O_5	O_6
O_1		6	5	1	5	1
O_2			1	5	2	7
O_3				5	1	6
O_4					4	2
O_5						6
O_6						

   a) 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.
   b) Using the obtained dendrogram, define an induced metrics and give a distance matrix for the considered set of objects.

3. **(15 p) Independent component analysis.** Describe the ICA method, paying attention to the following aspects.
    a) What is the input to this method? What is the output?
    b) Under which assumptions can ICA give good results? When will ICA fail to give the desired results?
    c) Which quantitative measure(s) is/are used to test if the method is giving good results and to check how near the solutions is?
    d) Name applications of ICA and describe what the input and the output to these applications are and what the goals of these applications are?
    e) What are the similarities and dissimilarities between ICA and PCA (principal component analysis)?
4. **(10 p) Clustering.** Describe the gap statistics method that is used to determine the number of clusters in a data set.
5. **(15 p) LVQ.** Describe the basic LVQ algorithm (LVQ1) and the relevance LVQ (RLVQ) algorithm? Formulate the rule for changing the prototypes and, in the case of RLVQ, also the rule for changing the relevances in an iteration. What methods can you name for empirically determining the test error of a classifier?
6. **(5 p) k-NN.** Justify the use of k-nearest neighbor classification using Bayesian theory.
7. **(10 p) Fingerprint matching.** Assume that you are given a set of 1000000 fingerprints. The set contains 1 fingerprint of the right hand ring finger of each of 1000000 persons. Assume that you have a program that can compare two fingerprints and give a similarity value.
    a) Describe how you would use this data and program to design an identification system based on statistical decision theory.
    b) Describe the main components of hypothesis testing in statistical decision theory. How would you apply it to fingerprint matching?