IN4085
Pattern Recognition

Written examination
23-06-2009, 14.00-17.00

- **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**

Answer sheet 1

Name :
Student number :

1 Statements *(10 points)*

Circle the correct statement, i.e. TRUE or FALSE. If a statement does not hold in general, but only under certain conditions that are not mentioned, then the statement should be marked as FALSE. 10 correct answers will give you 0 points; each additional correct answer gives you 1 point.

Classification

1. If no features are available, the optimal classifier is realized by assigning all objects to the class with the smallest prior probability.
   TRUE FALSE
2. The best classifier for a set of normally distributed classes can never be linear.
   TRUE FALSE
3. If the Mahalanobis distance between two classes is 0, then their distributions overlap entirely.
   TRUE FALSE
4. The maximum rule for combining classifiers selects the classifier that overall performs best.

   TRUE FALSE
5. Fisher's linear discriminant is scale insensitive.
   TRUE FALSE
6. The Parzen classifier (based on an optimal choice for the smoothing parameter) will for very large training sets approximate the Bayes classifier.
   TRUE FALSE
7. The support vector classifier is trained by minimizing the Bayes error.
   TRUE FALSE
8. The nearest mean classifier is always linear.
   TRUE FALSE
9. The logistic classifier maximizes the likelihood of the training set under the assumption of a linear classification boundary.
   TRUE FALSE
10. Classifiers based on a dissimilarity representation of the objects do not suffer from the curse of dimensionality.
    TRUE FALSE

Answer sheet 2

Name :
Student number :

Clustering

1. The k-means clustering algorithm assumes clusters can have any Gaussian distribution.
   TRUE FALSE

2. The Self-organizing map solution is scale independent.
   TRUE FALSE

3. k-means clustering results do not always reproduce exactly.
   TRUE FALSE

4. To compare two different clustering solutions, one can use the gap statistic.
   TRUE FALSE

5. The log-likelihood of probabilistic clustering models increases when the number of clusters is increased.
   TRUE FALSE

Answer sheet 3

Name :
Student number :

Feature Extraction and Selection

1. When classes are normally distributed with equal class covariance matrices, feature extraction by means of the Fisher criterion J_F always leads to the optimal lower-dimensional feature space (in terms of class overlap).
   TRUE FALSE

2. The selection of a subset of features can never reduce the intrinsic class overlap of the original problem.
   TRUE FALSE

3. In general, selecting 500 features out of 1,000 is computationally more demanding if you use feature backward selection then when using the forward selection approach.
   TRUE FALSE

4. Selecting three features is always better than selecting just two.
   TRUE FALSE

5. The number of different subspaces that feature extraction can find for a specific problem is *always* larger than the number of different subspaces feature selection can provide.
   TRUE FALSE

2 Classification *(10 points)*

We consider the suitability of the following classifiers:

NMC Nearest mean classifier

fisherc Fisher's linear discriminant

QDC Quadratic classifier assuming normal distributions

parzenc Parzen classifier with an optimized single smoothing parameter

SVC_1 Linear support vector machine

for the following two problems:

Problem_1 Two classes given in a 5-dimensional feature space by 1000 training examples each. It is known that the features are strongly correlated.

Problem_2 Two classes given in 256-dimensional feature space by 200 training examples each.

Answer the following questions:

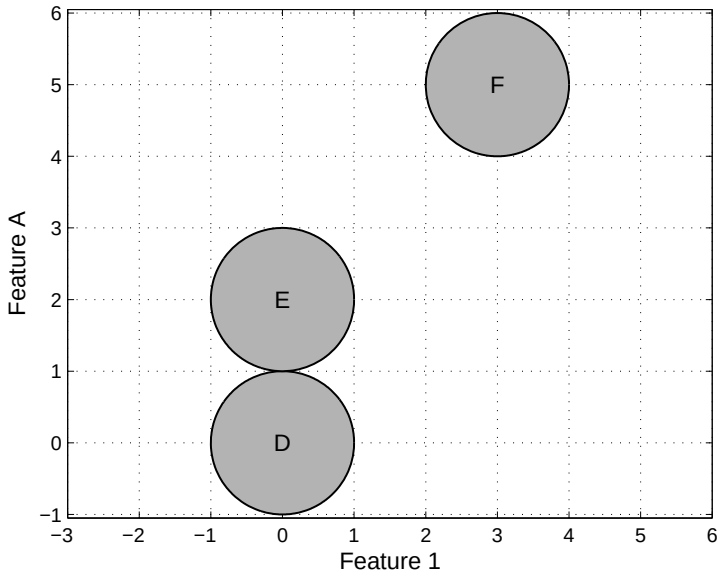
a. Suppose an advice has to be given for the direct use (without any further data investigations) of a classifier for Problem_1. What is your order of preference for the five classifiers. Why? *(3 points)*

b. Suppose an advice has to be given for the direct use (without any further data investigations) of a classifier for Problem_2. What is your order of preference for the five classifiers. Why? *(3 points)*

c. Give three options to improve the result for Problem_1 if all types of data-analysis are allowed, but just these 5 classifiers are available. Options may be combined. *(2 points)*

d. Give three options to improve the result for Problem_1 if all types of data-analysis are allowed, but just these 5 classifiers are available. Options may be combined. *(2 points)*

3 Feature Extraction and Selection *(10 points)*



Consider the three equally probable classes D, E, and F in a 2-dimensional feature space. See the figure above. All have a uniform and circular distributions of diameter 2 and the three class means are located in $(0, 0)$, $(0, 2)$, and $(3, 5)$, respectively. In addition, the distributions are such that all class covariance matrices, as well as the within-class covariance matrix $\mathbf{S}_W$, are equal to the identity matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

a. In a forward selection scheme, what feature is selected when the selection criterion used is the sum of the (squared) Mahalanobis distances? (You may also use the multi-class Fisher criterion J_{F} here as this leads to the same result.) *(1 points)*

b. Using the same criterion, what feature is selected when a branch-and-bound approach is used? *(1 points)*

c. Use the given class means to determine the between-class scatter, or covariance matrix. *(2 points)*

d. The two eigenvectors of the covariance matrix in the previous item c. are $(3, 2)$ and $(-2, 3)$. The corresponding eigenvalue of $(3, 2)$ is larger than the eigenvalue associated to $(-2, 3)$.

   Now, use the necessary information from the between-class covariance matrix, the within-class covariance matrix, the two eigenvectors, and the eigenvalue ordering to determine the 1-dimensional optimal solution based on KL5.

   N.B. KL5 is the fifth variant of the Karhunen-Loève transformation (see Webb) and is equal to the feature extraction based on the maximization of the Fisher criterion J_{F} (`fisherm` in terms of PRTools). *(2 points)*

e. Which of the three previous solutions (from a., b., and d.) performs worst? How much class overlap is attained when relying on the feature forward selection procedure? *(2 points)*

f. Change one or more of the current three class means and modify the problem such that the feature extraction performs better (i.e., gives lower class overlap) than the two feature

selection methods. Stated differently: provide three class means and show that the feature extraction approach above now outperforms the feature selection methods. *(2 points)*

4 Clustering *(10 points)*

Answer sheet 4(a) shows a dataset $\mathbf{X}$ with eight samples $x_1, \ldots x_8$.

a. Perform hierarchical clustering on the eight samples depicted in Answer sheet 4 using *complete* linkage and the *city block* distance as dissimilarity measure. The city block distance between two samples is measured exclusively along vertical and horizontal lines. For example, $d(x_2, x_3) = 5$ (horizontal) + 2 (vertical) = 7.
Draw the dendrogram in Answer sheet 4(b) and name the axes. *(3 points)*

b. Draw the fusion graph for the complete linkage clustering in Answer sheet 4(c) and name the axes. *(2 points)*

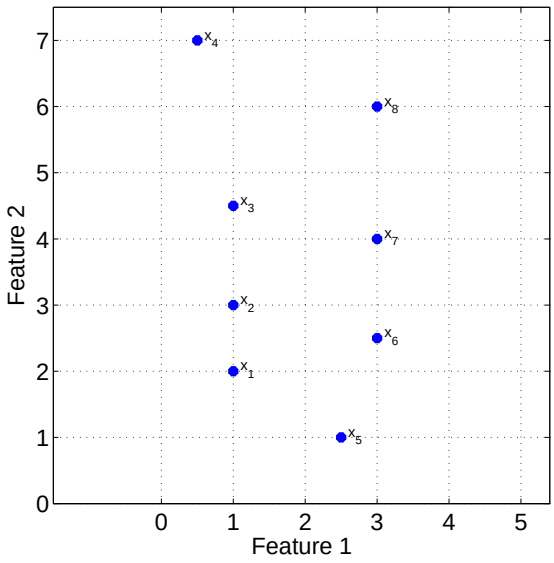
c. Given the fusion graph that you found in Answer 4(b), what is the natural number of clusters? Explain your answer. *(1 points)*

d. This clustering is not unique. Explain what type of solution the *single* linkage hierarchical clustering will give, and why this solution is obtained. *(2 points)*

e. Give at least one other element (next to the clustering method) that can be adapted to change the clustering result on the data. *(2 points)*

Answer sheet 4

Name :
Student number :



(a) Dataset

(b) Dendrogram, complete linkage

(c) Fusion graph, complete linkage