IN4085
Pattern Recognition

Written examination
18-01-2006, 14.00-17.00

- **There are four questions**
- **Answer each question on a separate sheet**
- **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. Forward feature selection is slower than individual selection.
   TRUE FALSE

2. Fisher's linear discriminant is scale insensitive.
   TRUE FALSE

3. In general, more biased classifiers will also have higher variance.
   TRUE FALSE

4. To apply multidimensional scaling (MDS), it is not necessary to measure features on objects.
   TRUE FALSE

5. For a finite training set, the average cross-validation error will be smaller than the true error.
   TRUE FALSE

6. Sphering rescales the original features to a new set of unit standard deviation features.
   TRUE FALSE

7. The support vector classifier is trained by minimizing the Bayes error.
   TRUE FALSE

8. If there are insufficient samples to estimate per-class covariance matrices well, one should use the quadratic Bayes plug-in classifier.
   TRUE FALSE

9. The nearest mean classifier is always linear.
   TRUE FALSE

10. If a non-support vector is removed from the training set and the support vector classifier is trained again, the classifier will not change.
    TRUE FALSE

Answer sheet 2

Name :
Student number :

Regression

1. Artificial neural networks with sigmoid transfer functions in the hidden and output units cannot approximate any real-valued function.
   TRUE FALSE

2. Maximum *a posteriori* (MAP) estimation does not take the *a priori* probability of a certain solution into account.
   TRUE FALSE

3. Local weighted linear regression suffers less from boundary bias than nearest neighbour regression.
   TRUE FALSE

4. When the sum-of-squares due to linear regression is small compared to the sum-of-squares about the mean, the fit is good.
   TRUE FALSE

5. A low R^2 for linear regression indicates that choosing a nonlinear model will drastically improve results.
   TRUE FALSE

Answer sheet 3

Name :
Student number :

Clustering

1. k-means clustering can be seen as a special case of the EM algorithm, in which all Gaussians are spherical and identical.
   TRUE FALSE

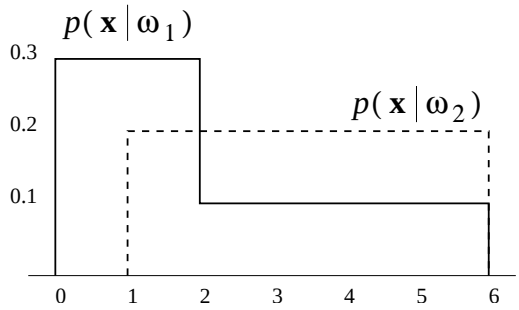
2. Single linkage gives more compact clusters than complete linkage.
   TRUE FALSE

3. Hierarchical clustering results depend on the initialization.
   TRUE FALSE

4. In hierarchical clustering, the largest jump in the fusion graph indicates where the dendogram should best be cut.
   TRUE FALSE

5. Self-organizing maps perform both clustering and feature extraction.
   TRUE FALSE

2 Classification *(10 points)*



The above figure shows the class-conditional probability density functions for two classes.

a. Compute the Bayes error when $p(\omega_1) = 0.5$, $p(\omega_2) = 0.5$. *(2 points)*

b. Compute the Bayes error when $p(\omega_1) = 0.2$, $p(\omega_2) = 0.8$. *(2 points)*

c. Given that we use the Bayes classifier of a. and given the cost matrix

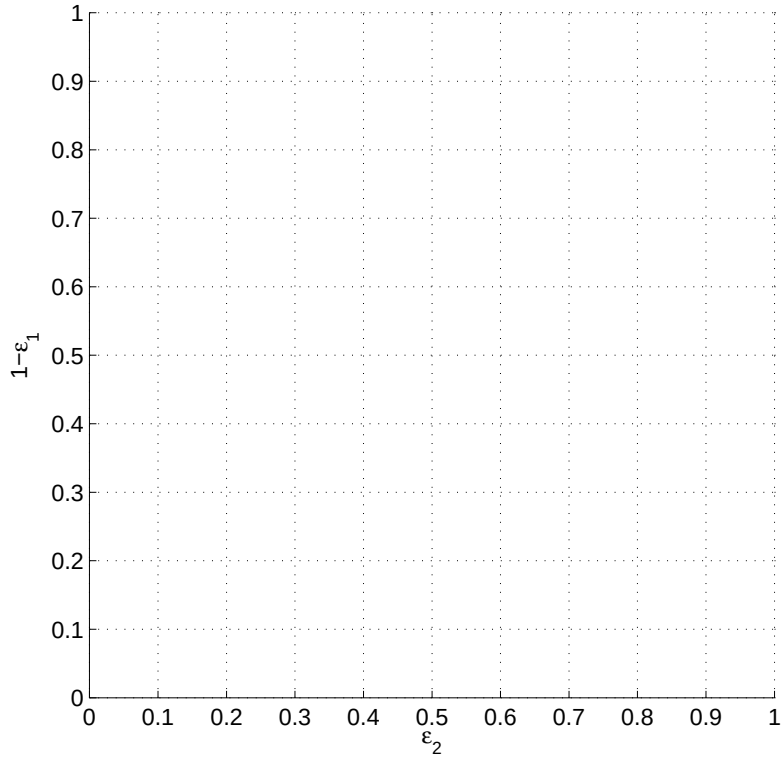
$$C = \begin{pmatrix} 0 & 4 \\ 1 & 0 \end{pmatrix},$$

compute the expected cost for the case that $p(\omega_1) = 0.5$, $p(\omega_2) = 0.5$. *(3 points)*

d. Draw the receiver-operator curve (ROC) for the Bayes classifier, in Answer sheet 4. *(3 points)*

Answer sheet 4

Name :
Student number :



Note: ϵ_1 indicates the error made on objects from class 1, ϵ_2 the error on objects from class 2.

3 Clustering *(10 points)*

Answer sheet 5(a) shows a dataset $\mathbf{X}$ with seven samples $x_1, \ldots x_7$.

a. Perform hierarchical clustering on the seven samples depicted in Answer sheet 5 using *single* 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 5(b). *(3 points)*

b. Draw the fusion graph for the single linkage clustering you found above, in Answer sheet 5(c). *(1 points)*

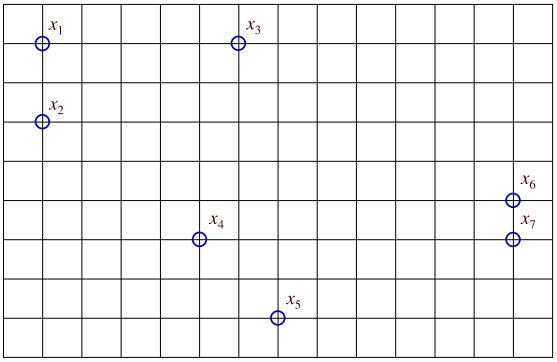
c. Now perform hierarchical clustering on the seven samples depicted in Answer sheet 5(a) using *complete* linkage and the *city block* distance as dissimilarity measure.
Draw the dendrogram in Answer sheet 5(d). *(3 points)*

d. Draw the fusion graph for the complete linkage clustering you found above, in Answer sheet 5(e). *(1 points)*

e. Based on the two fusion graphs, which clustering do you think is the best? Motivate your answer. *(2 points)*

Answer sheet 5

Name :
Student number :



(a) Dataset

(b) Dendrogram, single linkage

(c) Fusion graph, single linkage

(d) Dendrogram, complete linkage

(e) Fusion graph, complete linkage

4 Regression *(10 points)*

Answer sheet 6 shows a dataset with seven inputs x and corresponding outputs y (indicated by blue circles).

a. Draw the 1-nearest neighbour smoothing estimate $\hat{y}$ over the entire range of x, in Answer sheet 6(a). *(3 points)*

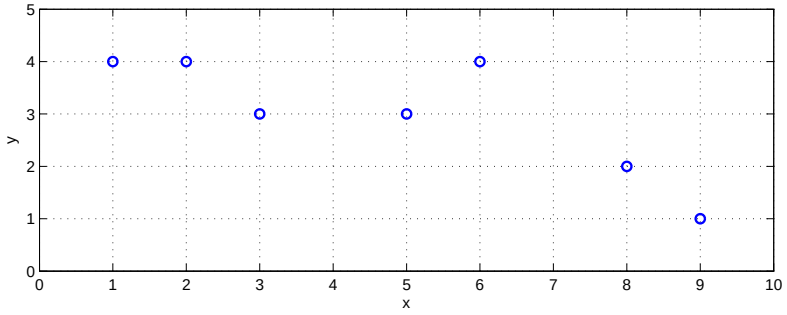
b. Draw the 3-nearest neighbour smoothing estimate $\hat{y}$ over the entire range of x, in Answer sheet 6(b). *(2 points)*

c. Draw the Gaussian kernel smoothing estimate $\hat{y}$ for $\sigma \to \infty$ over the entire range of x, in Answer sheet 6(c). *(3 points)*

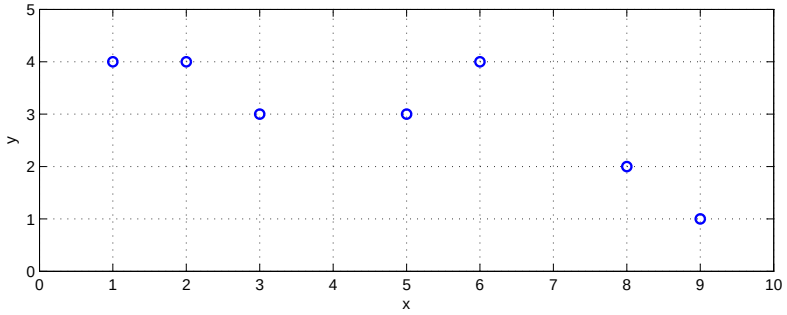
d. Draw the Gaussian kernel smoothing estimate $\hat{y}$ for $\sigma \to 0$ over the entire range of x, in Answer sheet 6(d). *(2 points)*

Answer sheet 6

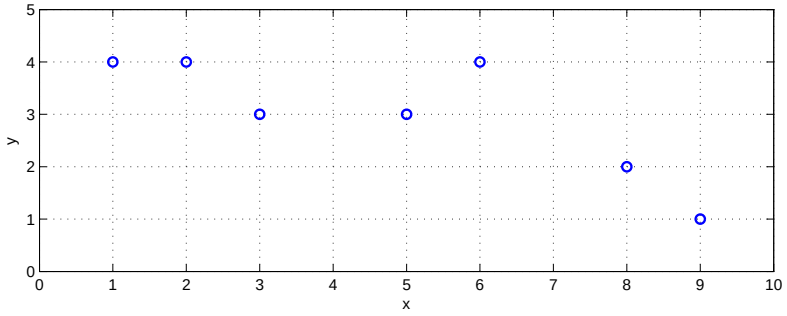
Name :
Student number :



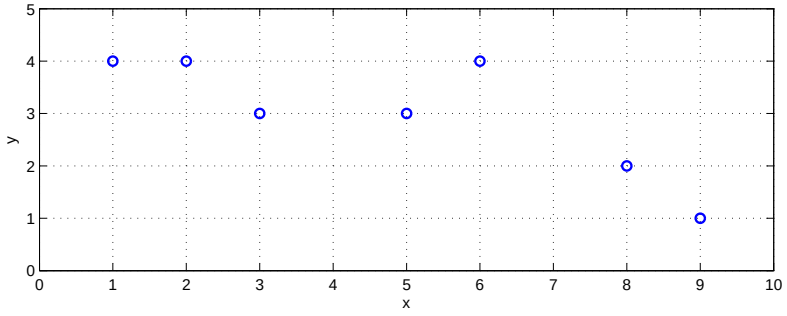
(a) 1-NN smoother



(b) 3-NN smoother



(c) Gaussian kernel smoother ($\sigma \to \infty$)



(d) Gaussian kernel smoother ($\sigma \to 0$)