Univ.-Prof. Dr. rer. nat. Rudolf Mathar

1	2	3	4	5	6	7	8	$\sum$
13	12	14	15	15	13	12	6	100

Written Examination

Fundamentals of Big Data Analytics

Monday, March 12, 2018, 02:00 p.m.

Name: ______________________________ Matr.-No.: ______________

Field of study: ____

Please pay attention to the following:

1) The exam consists of **8 problems**. Please check the completeness of your copy. **Only** written solutions on these sheets will be considered. Removing the staples is **not** allowed.

2) The exam is passed with at least **50 points**.

3) You are free in choosing the order of working on the problems. Your solution shall clearly show the approach and intermediate arguments.

4) **Admitted materials:** The sheets handed out with the exam and a non-programmable calculator.

5) The results will be published on Friday evening, the 16.03.18, on the homepage of the institute.

The corrected exams can be inspected on Friday, 23.03.18, 10:00h. at the seminar room 333 of the Chair for Theoretical Information Technology, Kopernikusstr. 16.

Acknowledged: ______________________________

(Signature)

Problem 1. (13 points)
Maximum Likelihood Estimator:

The Burr Distribution is commonly used to model household income. Its cumulative distribution function is given by

$$F(x|\theta) = \begin{cases} 0, & x < 0 \\ 1 - \frac{1}{(1+x^2)^{1/\theta}}, & x \geq 0 \end{cases}$$

where $\theta > 0$. Assume i.i.d. samples $\mathbf{X} = X_1, X_2, \ldots, X_n$ are taken from the Burr distribution, and let $\mathbf{X} = [X_1, X_2, \ldots, X_n]^{\mathrm{T}}$.

a) Find the probability density function of the Burr distribution. (2P)

b) Find the log likelihood function $\ell(\mathbf{X}; \theta)$ of $\mathbf{X}$. (4P)

c) Find the maximum likelihood estimator (MLE) $\hat{\theta}$ of θ based on $\mathbf{X}$. (4P)

d) Is the above MLE estimator unbiased? Justify your answer.
Hint: Use without verifying that for all X_i, $\mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}\theta} \ln f(X_i|\theta)\right] = 0$. (2P)

Problem 2. (12 points)
Principal Component Analysis (PCA):

a) Let $\mathbf{A}$ be a symmetric $n \times n$ matrix. Show that there exists a real $t > 0$, large enough such that $\mathbf{A} + t\mathbf{I}$ is positive definite. What is the minimum value of t? (4P)

Assume that $\mathbf{A}$ is given by:

$$\mathbf{A} = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} \begin{pmatrix} 2 & 2 & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \end{pmatrix}$$

b) What is the rank of $\mathbf{A}$? (1P)

c) Calculate the spectral decomposition $\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}$ of $\mathbf{A}$ by determining the matrices $\mathbf{V}$ and $\mathbf{\Lambda}$. (4P)

d) Assume that $\frac{1}{3}\mathbf{A}$ is a sample covariance matrix. Determine the projection matrix $\mathbf{Q}$ for PCA to transform three-dimensional samples to two dimensions. (2P)

e) Determine the projection error $\frac{1}{n-1} \max_{\mathbf{Q}} \sum_{i=1}^{n} ||\mathbf{Q}\mathbf{x}_i - \mathbf{Q}\bar{\mathbf{x}}_n||^2$ for the above choice of $\mathbf{Q}$. (2P)

Problem 3. (14 points)
Diffusion Map:

The dataset shown in Figure 1 is composed of 8 points $\mathbf{x}_1, \dots, \mathbf{x}_8 \in \mathbb{R}^2$.

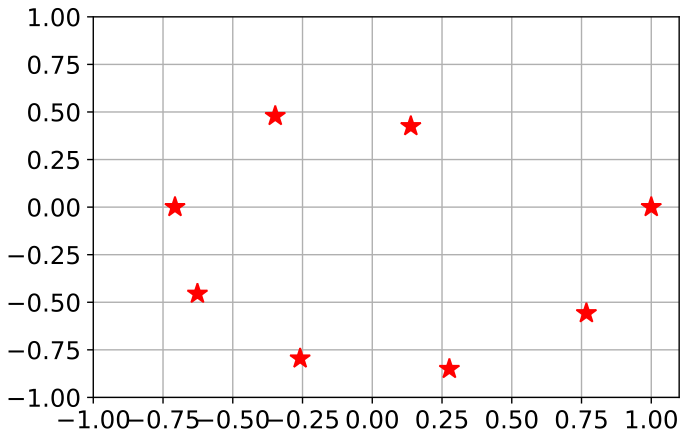


Figure 1: Data Points

The following matrix $\mathbf{\Delta} \in \mathbb{R}^{8\times 8}$ is the Euclidean distance matrix for these points.

$$\mathbf{\Delta} = \begin{pmatrix} 0.0 & 0.2 & 0.9 & 1.4 & 1.6 & 1.6 & 1.4 & 0.9 \\ 0.2 & 0.0 & 0.4 & 0.9 & 1.6 & 2.2 & 2.3 & 2.0 \\ 0.9 & 0.4 & 0.0 & 0.2 & 0.8 & 1.7 & 2.5 & 2.9 \\ 1.4 & 0.9 & 0.2 & 0.0 & 0.3 & 1.0 & 2.0 & 2.9 \\ 1.6 & 1.6 & 0.8 & 0.3 & 0.0 & 0.3 & 1.1 & 2.2 \\ 1.6 & 2.2 & 1.7 & 1.0 & 0.3 & 0.0 & 0.3 & 1.2 \\ 1.4 & 2.3 & 2.5 & 2.0 & 1.1 & 0.3 & 0.0 & 0.4 \\ 0.9 & 2.0 & 2.9 & 2.9 & 2.2 & 1.2 & 0.4 & 0.0 \end{pmatrix}$$

Assume that we want to construct a diffusion map using the following kernel function:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} \exp(-5\|\mathbf{x}_j - \mathbf{x}_i\|_2^2)\,, & \|\mathbf{x}_j - \mathbf{x}_i\|_2 \leq 0.8\,, \\ 0\,, & \text{otherwise} \end{cases} .$$

Using this kernel function calculate

a) the weight matrix $\mathbf{W}$ for the difussion map, (4P)

b) the first 2 rows of the transition matrix $\mathbf{M}$ for the difussion map. (3P)

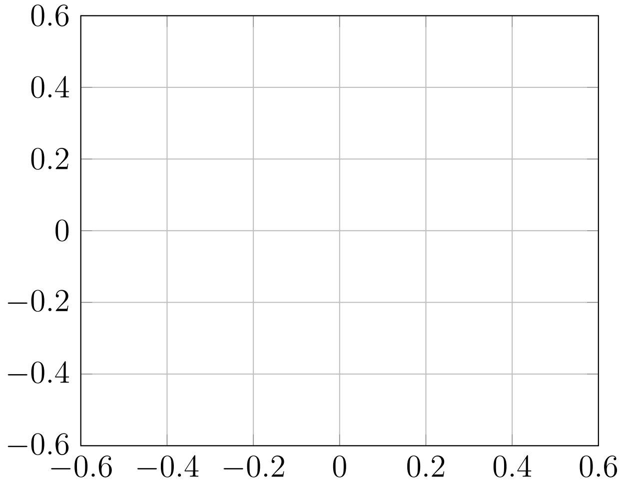
The spectral decomposition of $\mathbf{S} = \mathbf{D}^{\frac{1}{2}}\mathbf{M}\mathbf{D}^{-\frac{1}{2}}$, with $\mathbf{D} = \operatorname{diag}(\deg(1), \dots, \deg(8))$, is given by $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$, where $\mathbf{V}$ contains the eigenvectors, and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_8)$ the eigenvalues.

Suppose that

$$\mathbf{D}^{-\frac{1}{2}}\mathbf{V} = \begin{pmatrix} 0.2 & 0.3 & 0.3 & -0.2 & -0.2 & -0.3 & 0.3 & 0.1 \\ 0.2 & 0.3 & 0.2 & -0. & 0.1 & 0.3 & -0.3 & -0.2 \\ 0.2 & 0.2 & -0.2 & 0.3 & 0.3 & 0.2 & 0.2 & 0.3 \\ 0.2 & 0.1 & -0.3 & 0.2 & -0.1 & -0.3 & -0. & -0.4 \\ 0.2 & -0.1 & -0.3 & -0.2 & -0.4 & 0.1 & -0.2 & 0.3 \\ 0.2 & -0.2 & -0.1 & -0.3 & 0.1 & 0.2 & 0.3 & -0.2 \\ 0.2 & -0.3 & 0.2 & -0.1 & 0.3 & -0.3 & -0.3 & 0.1 \\ 0.2 & -0.4 & 0.4 & 0.4 & -0.3 & 0.2 & 0.1 & -0.1 \end{pmatrix}$$

and $\mathbf{\Lambda} = \text{diag}([1.0,\ 0.95,\ 0.83,\ 0.65,\ 0.39,\ 0.15,\ 0.02,\ -0.1])$.

c) Calculate and draw the truncated difussion maps $\phi_t^{(2)}(\mathbf{x}_i)$ for $i = 1, \ldots, 8$ and $t = 0$. (5P)



d) Explain what happens to the truncated diffusion maps $\phi_t^{(2)}(\mathbf{x}_i)$ as $t \to \infty$. (2P)

Problem 4. (15 points)
Discriminant Analysis:

A training dataset consists of 4 vectors $\mathbf{x}_1, \ldots, \mathbf{x}_4 \in \mathbb{R}^2$ belonging to two classes C_1 and C_2. The vectors are given by

$$\mathbf{x}_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \text{and} \quad \mathbf{x}_4 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Suppose that $\mathbf{x}_1, \mathbf{x}_2$ and $\mathbf{x}_3$ belong to C_1, and $\mathbf{x}_4$ belongs to C_2 as shown in Figure 2 .

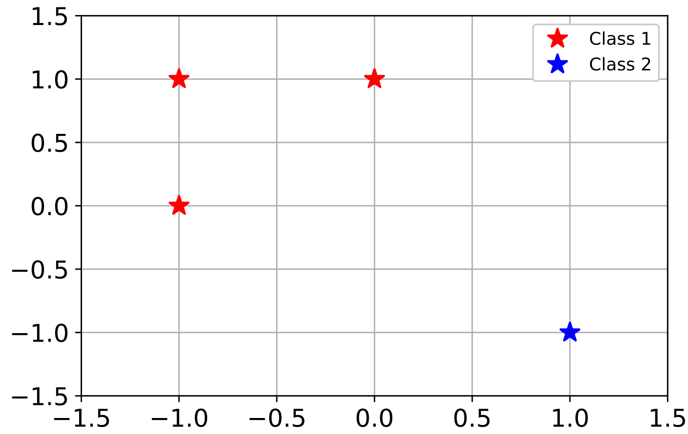


Figure 2: Data Points

The discriminant vector $\mathbf{a} \in \mathbb{R}^2$ is given as $\mathbf{a} = \frac{1}{\sqrt{2}}(-1, 1)^T$.

a) The separating hyperplane has the form $\mathbf{a}^T\mathbf{x} - b = 0$. Calculate the value of $b \in \mathbb{R}$ and draw the separating hyperplane on Figure 2 . (4P)

b) Calculate the *sum of squares between groups.* (3P)

c) Calculate the *sum of squares within groups.* (4P)

Assume that $\tilde{\mathbf{x}}_4 \in \mathbb{R}^2$ is a noisy version of $\mathbf{x}_4$ such that

$$\tilde{\mathbf{x}}_4 = \mathbf{x}_4 + \epsilon \boldsymbol{\eta},$$

where $\boldsymbol{\eta} \in \mathbb{R}^2$ is a vector with $\|\boldsymbol{\eta}\|_2 = 1$ and $\epsilon > 0$.

d) Find the minimum ϵ such that $\tilde{\mathbf{x}}_4$ gets allocated to C_1 by the discriminant rule. (4P)

Problem 5. (15 points)
Support Vector Machines:

A training dataset is composed of six vectors $\mathbf{x}_i$ in $\mathbb{R}^2$, $i = 1, \ldots, 6$, belonging to two classes. The class membership is indicated by the labels $y_i \in \{-1, +1\}$. A kernel-based support vector machine is used to find the maximum-margin hyperplane by solving the following dual problem:

$$\begin{aligned} \max_{\boldsymbol{\lambda}} \quad & \sum_{i=1}^{6} \lambda_i - \frac{1}{2} \sum_{i=1}^{6} \sum_{j=1}^{6} y_i y_j \lambda_i \lambda_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{s.t.} \quad & 0 \leq \lambda_i \leq 2 \qquad \text{and} \qquad \sum_{i=1}^{6} \lambda_i y_i = 0. \end{aligned}$$

The kernel function is given by:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|_2^2).$$

The value of γ is chosen as 0.6.

The dataset and the outputs of the optimization problem are given in the following table.

Data	Label	Solution	Data	Label	Solution
$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$y_1 = -1$	$\lambda_1^\star = 2$	$\mathbf{x}_4 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$	$y_4 = 1$	$\lambda_4^\star = 2$
$\mathbf{x}_2 = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$	$y_2 = -1$	$\lambda_2^\star = 0.74$	$\mathbf{x}_5 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$	$y_5 = 1$	$\lambda_5^\star = 0.5$
$\mathbf{x}_3 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$	$y_3 = -1$	$\lambda_3^\star = 1.76$	$\mathbf{x}_6 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$	$y_6 = 1$	$\lambda_6^\star = 2$

a) Determine the support vectors. (6P)

b) Determine the kernel-based classifier by specifying all the parameters. (6P)
Hint: Round the numbers to the nearest thousandth, e.g., $0.0014 \approx 0.001$ or $0.0016 \approx 0.002$ or $0.0015 \approx 0.002$.

c) Suppose that γ is very large so that the kernel function can be approximated by

$$K(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 1 & \mathbf{x}_i = \mathbf{x}_j \\ 0 & \text{otherwise} \end{cases}.$$

Determine the support vectors for this problem. (3P)

Problem 6. (13 points)
Kernels for SVM:

a) Determine the following kernel functions are valid kernels for support vector machines and explain the reason. (6P)

  a) $K(\mathbf{x}_i, \mathbf{x}_j) = 1$ for all $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^p$.

  b) $K(\mathbf{x}_i, \mathbf{x}_j) = \max_{k \in \{1,\ldots,p\}} (x_i(k) - x_j(k))$ for $\mathbf{x}_i = (x_i(1), \ldots, x_i(p))^T$ and $\mathbf{x}_j = (x_j(1), \ldots, x_j(p))^T$.

  c) $K(\mathbf{x}_i, \mathbf{x}_j) = |\|\mathbf{x}_i\|_2^2 - \|\mathbf{x}_j\|_2^2|$ for all $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^p$.

b) Suppose that a kernel is given by $K(\mathbf{x}, \mathbf{y}) = 4(\mathbf{x}^T\mathbf{y})^2 + 3(\mathbf{x}^T\mathbf{y}) + 1$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$. Find the feature function for this kernel. Determine the dimension of the feature space. (7P)

Problem 7. (12 points)
Clustering:
Part I
The set $\Phi = \{\mathbf{x}_i \mid i = 1, \ldots, 6\} \subset \mathbb{R}^2$, with

$$\mathbf{x}_1 = \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \ \mathbf{x}_2 = \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \ \mathbf{x}_3 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ \mathbf{x}_4 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \ \mathbf{x}_5 = \begin{pmatrix} -2 \\ -2 \end{pmatrix}, \ \mathbf{x}_6 = \begin{pmatrix} -1 \\ -4 \end{pmatrix}.$$

a) The k-means clustering algorithm is used to partition Φ into two clusters: C_1 and C_2. At a certain iteration, $\mathbf{x}_1$ and $\mathbf{x}_5$ are the center of C_1 and C_2, respectively. Assign each sample in Φ to the appropriate clusters. Suppose the Euclidian distance is used for the assignment. (4P)

b) Determine the centers of the two clusters according to the update in **a)**. (2P)

Part II
The table below shows the pairwise dissimilarities between four points in a dataset Γ, where $\Gamma = \{P_1, P_2, P_3, P_4, P_5\}$.

	P_1	P_2	P_3	P_4	P_5
P_1	0	0.9	0.8	0.3	0.4
P_2	0.9	0	0.5	0.2	0.1
P_3	0.8	0.5	0	0.6	0.2
P_4	0.3	0.2	0.6	0	0.7
P_5	0.4	0.1	0.2	0.7	0

Use the agglomerative clustering algorithm to partition Γ into two clusters: C_1 and C_2. For this assignment, use the average linkage distance between C_1 and C_2, which is given by

$$d(C_1, C_2) = \frac{1}{\mid C_1 \mid\mid C_2 \mid} \sum_{i \in C_1, j \in C_2} \delta_{i,j},$$

where $\mid . \mid$ denotes the cardinality, and $\delta_{i,j}$ is the dissimilarity between points i and j. (6P)

Problem 8. (6 points)
Regression:
Assume the signal-to-noise ratio (SNR) in dB at a certain receiver is indicated by the variable $x \in \mathbb{R}$. The receiver should have a bit error rate (BER) below a certain threshold, so that the message is decodable. The variable $y \in \{0, 1\}$ models this information; $y = 0$ indicates a decodable message, and $y = 1$ indicates a non-decodable message. Assume logistic regression is used to model y as a function of x.

a) At a given iteration, $\nu = (\nu_0, \nu_1)$ denotes the estimated coefficients of the model, which are given by $(\nu_0, \nu_1) = (-0.05, 0.08)$. Assume the sigmoid function is used as a non-linear function in logistic regression. Estimate the probability that a message is decodable at $x = 10$ dB. (4P)

b) Repeat **a)** using the following activation function (2P)

$$f(x) = \log_{10}(1 + \exp(x)).$$

Additional sheet

Problem:

Additional sheet

Problem:

Additional sheet

Problem:

Additional sheet

Problem:

Additional sheet

Problem: