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 \( X = X_1, X_2, \ldots, X_n \) are taken from the Burr distribution, and let \( X = [X_1, X_2, \ldots, X_n]^T \).

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

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

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

d) Is the above MLE estimator unbiased? Justify your answer.

Hint: Use without verifying that for all \( X_i \),
\[ \mathbb{E} \left[ \frac{d}{d\theta} \ln f(X_i|\theta) \right] = 0. \] (2P)
Problem 2. (12 points)

Principal Component Analysis (PCA):

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

Assume that $A$ is given by:

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

b) What is the rank of $A$? (1P)

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

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

e) Determine the projection error $\frac{1}{n-1} \max_Q \sum_{i=1}^n \|Qx_i - \bar{Q}x_n\|^2$ for the above choice of $Q$. (2P)
Problem 3. (14 points)

Diffusion Map:

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

![Figure 1: Data Points](image)

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

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

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 diffusion map, (4P)

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

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

Suppose that
\[ D^{-\frac{1}{2}}V = \begin{pmatrix}
0.2 & 0.3 & 0.3 & -0.2 & -0.2 & -0.3 & 0.3 & 0.1 \\
0.2 & 0.3 & 0.2 & -0.1 & 0.1 & 0.3 & -0.3 & -0.2 \\
0.2 & 0.2 & -0.2 & 0.3 & 0.3 & 0.2 & 0.2 & 0.3 \\
0.2 & 0.1 & -0.3 & 0.2 & -0.1 & -0.3 & -0.3 & -0.4 \\
0.2 & -0.1 & -0.3 & -0.2 & -0.4 & 0.1 & -0.2 & 0.3 \\
0.2 & -0.2 & -0.1 & -0.3 & 0.1 & 0.2 & 0.3 & -0.2 \\
0.2 & -0.3 & 0.2 & -0.1 & 0.3 & -0.3 & -0.3 & 0.1 \\
0.2 & -0.4 & 0.4 & 0.4 & -0.3 & 0.2 & 0.1 & -0.1 
\end{pmatrix} \]

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

c) Calculate and draw the truncated diffusion maps \( \phi^{(2)}_t(x_i) \) for \( i = 1, \ldots, 8 \) and \( t = 0 \). (5P)

![Graph](image)

d) Explain what happens to the truncated diffusion maps \( \phi^{(2)}_t(x_i) \) as \( t \to \infty \). (2P)
Problem 4. (15 points)

**Discriminant Analysis:**

A training dataset consists of 4 vectors \( x_1, \ldots, x_4 \in \mathbb{R}^2 \) belonging to two classes \( C_1 \) and \( C_2 \). The vectors are given by

\[
x_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad x_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \text{and} \quad x_4 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.
\]

Suppose that \( x_1, x_2 \) and \( x_3 \) belong to \( C_1 \), and \( x_4 \) belongs to \( C_2 \) as shown in Figure 2.

![Figure 2: Data Points](image)

The discriminant vector \( a \in \mathbb{R}^2 \) is given as

\[
a = \frac{1}{\sqrt{2}} (-1, 1)^T.
\]

**a)** The separating hyperplane has the form \( a^T x - b = 0 \). 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 \( \tilde{x}_4 \in \mathbb{R}^2 \) is a noisy version of \( x_4 \) such that

\[
\tilde{x}_4 = x_4 + \epsilon \eta,
\]

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

**d)** Find the minimum \( \epsilon \) such that \( \tilde{x}_4 \) 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 \( x_i \) 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:

\[
\begin{align*}
\max_{\lambda} \quad & \sum_{i=1}^{6} \lambda_i - \frac{1}{2} \sum_{i=1}^{6} \sum_{j=1}^{6} y_i y_j \lambda_i \lambda_j K(x_i, x_j) \\
\text{s.t.} \quad & 0 \leq \lambda_i \leq 2 \quad \text{and} \quad \sum_{i=1}^{6} \lambda_i y_i = 0.
\end{align*}
\]

The kernel function is given by:

\[
K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|_2^2).
\]

The value of \( \gamma \) is chosen as 0.6.

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

<table>
<thead>
<tr>
<th>Data</th>
<th>Label</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 = \begin{pmatrix} 1 \ 1 \end{pmatrix} )</td>
<td>( y_1 = -1 )</td>
<td>( \lambda_1^* = 2 )</td>
</tr>
<tr>
<td>( x_2 = \begin{pmatrix} -2 \ -1 \end{pmatrix} )</td>
<td>( y_2 = -1 )</td>
<td>( \lambda_2^* = 0.74 )</td>
</tr>
<tr>
<td>( x_3 = \begin{pmatrix} -1 \ -1 \end{pmatrix} )</td>
<td>( y_3 = -1 )</td>
<td>( \lambda_3^* = 1.76 )</td>
</tr>
<tr>
<td>( x_4 = \begin{pmatrix} -1 \ 0 \end{pmatrix} )</td>
<td>( y_4 = 1 )</td>
<td>( \lambda_4^* = 2 )</td>
</tr>
<tr>
<td>( x_5 = \begin{pmatrix} -2 \ 1 \end{pmatrix} )</td>
<td>( y_5 = 1 )</td>
<td>( \lambda_5^* = 0.5 )</td>
</tr>
<tr>
<td>( x_6 = \begin{pmatrix} 1 \ 2 \end{pmatrix} )</td>
<td>( y_6 = 1 )</td>
<td>( \lambda_6^* = 2 )</td>
</tr>
</tbody>
</table>

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 \( \gamma \) is very large so that the kernel function can be approximated by

\[
K(x_i, x_j) = \begin{cases} 
1 & \text{if } x_i = x_j \\
0 & \text{otherwise}
\end{cases}
\]

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) \( K(x_i, x_j) = 1 \) for all \( x_i, x_j \in \mathbb{R}^p \).

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

c) \( K(x_i, x_j) = ||x_i||^2_2 - ||x_j||^2_2 \) for all \( x_i, x_j \in \mathbb{R}^p \).

b) Suppose that a kernel is given by \( K(x, y) = 4(x^T y)^2 + 3(x^T y) + 1 \) for \( x, y \in \mathbb{R}^p \). Find the feature function for this kernel. Determine the dimension of the feature space. (7P)
Problem 7. (12 points)

Clustering:

Part I
The set $\Phi = \{x_i \mid i = 1, \ldots, 6\} \subset \mathbb{R}^2$, with

$$x_1 = \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad x_4 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad x_5 = \begin{pmatrix} -2 \\ -2 \end{pmatrix}, \quad x_6 = \begin{pmatrix} -1 \\ -4 \end{pmatrix}.$$

a) The $k$-means clustering algorithm is used to partition $\Phi$ into two clusters: $C_1$ and $C_2$. At a certain iteration, $x_1$ and $x_5$ are the center of $C_1$ and $C_2$, respectively. Assign each sample in $\Phi$ 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 $\Gamma$, where $\Gamma = \{P_1, P_2, P_3, P_4, P_5\}$.

<table>
<thead>
<tr>
<th></th>
<th>P_1</th>
<th>P_2</th>
<th>P_3</th>
<th>P_4</th>
<th>P_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_1</td>
<td>0</td>
<td>0.9</td>
<td>0.8</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>P_2</td>
<td>0.9</td>
<td>0</td>
<td>0.5</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>P_3</td>
<td>0.8</td>
<td>0.5</td>
<td>0</td>
<td>0.6</td>
<td>0.2</td>
</tr>
<tr>
<td>P_4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.6</td>
<td>0</td>
<td>0.7</td>
</tr>
<tr>
<td>P_5</td>
<td>0.4</td>
<td>0.1</td>
<td>0.2</td>
<td>0.7</td>
<td>0</td>
</tr>
</tbody>
</table>

Use the agglomerative clustering algorithm to partition $\Gamma$ 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

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

where $| . |$ 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: