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

1	2	3	4	5	6	7	8	Σ
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:

a)

$$f(x|\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} F(x|\theta) = \frac{2x}{\theta(1+x^2)^{1/\theta+1}}$$

b)

$$\ell_i(\theta) = \ln f(x_i|\theta) = \ln 2x_i - \ln\theta - (\frac{1}{\theta}+1)\ln(1+x_i^2)$$

$$\ell(\theta) = \sum_{i=1}^{n} l_i(\theta) = -n\ln + \theta \sum_{i=1}^{n} \ln 2x_i - (\frac{1}{\theta}+1)\ln(1+x_i^2)$$

c)

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \ln f(x|\theta) = -\frac{1}{\theta} + \frac{1}{\theta^2}\ln(1+x^2)$$

$\hat{\theta}$ satisfies $\sum_{i=1}^{n}(-\frac{1}{\theta} + \frac{1}{\theta^2}\ln(1+x_i^2) = 0$ this results in $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\ln(1+x_i^2)$

d)

$$\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta} \ln f(x|\theta) = \mathbb{E}(-\frac{1}{\theta} + \frac{1}{\theta^2}\ln(1+x^2)) = 0 \quad \mathbb{E}\ln(1+x^2) = \theta$$

$$\begin{aligned}
\mathbb{E}\hat{\theta} &= \mathbb{E}\frac{1}{n}\sum_{i=1}^{n}\ln(1+x_i^2) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\ln(1+x_i^2) \\
&= \frac{1}{n}n\theta = \theta
\end{aligned}$$

Problem 2. (12 points)
Principal Component Analysis:

a) Let $\mathbf{A}$ be a symmetric, $n \times n$, matrix. Show that there exists a real, positive t, large enough such that $\mathbf{A} + t\mathbf{I}$ is positive definite. What is the minimum value of t? (5P)
Since $\mathbf{A}$ is symmetric, $\mathbf{A} + t\mathbf{I}$ is also symmetric. For any eigenvalue λ of $\mathbf{A}$, $\lambda + t$ is an eigenvalue of $\mathbf{A} + t\mathbf{I}$ so if $t > -\min_i \lambda_i$ then $\mathbf{A} + t\mathbf{I}$ has all (real) eigenvalues greater than 0, thus is positive definite.

Now 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) 3

c) Calculate the spectral decomposition $\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}$ of $\mathbf{A}$ by determining the matrices $\mathbf{V}$ and $\mathbf{\Lambda}$. (6P)

$$\mathbf{A} = \begin{pmatrix} 6 & 4 & 0 \\ 4 & 6 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{vmatrix} 6-\lambda & 4 & 0 \\ 4 & 6-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{vmatrix} = (1-\lambda)[(6-\lambda)(6-\lambda)-16] = (1-\lambda)(\lambda^2-12\lambda-20) = (1-\lambda)(\lambda-2)(\lambda-10) = 0\,.$$

This results in $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 10$.
From the above construction of $\mathbf{A}$ and using $\mathbf{A}v = \lambda v$ we get that the corresponding eigenvectors are $v_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} v_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$. After normalization of these vectors and combining to make $\mathbf{V}$ and $\mathbf{\Lambda}$ we get

$$\mathbf{V} = \begin{pmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \end{pmatrix}$$

and

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 10 \end{pmatrix}$$

d) Determine the best projection matrix $\mathbf{Q}$ to transform the three-dimensional samples to two-dimensions. (2P)
The best projection matrix $\mathbf{Q}$ is determined by the first k dominant eigenvectors $\mathbf{v}_i$ as $\mathbf{Q} \sum_{i=1}^{k} \mathbf{v}_i \mathbf{v}_i^{\mathrm{T}}$, where k is the dimension of the image. For a transformation of a three-dimensional sample to a two-dimensional data (k=2), we obtain

$$\mathbf{Q} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} + \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

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

The residuum $\frac{1}{n-1} \max_{\mathbf{Q}} \sum_{i=1}^{n} ||\mathbf{Q}\mathbf{x}_i - \mathbf{Q}\bar{\mathbf{x}}_n||^2$ is equal to the sum $\sum_{i=1}^{k} \lambda(\mathbf{S})$ of dominant eigenvalues, that is equal to $10/3 + 2/3 = 12/3 = 4$.

Problem 3. (14 points)
Diffusion Map, (12P):

a) Since most of the euclidean distances are greater or equal than 0.8, most of the values in $\mathbf{W}$ are equal to zero. Then, we only need to calculate $e^{-5\cdot(0.2)^2} = 0.82$, $e^{-5\cdot(0.3)^2} = 0.64$, $e^{-5\cdot(0.4)^2} = 0.45$, thus

$$\mathbf{W} = \begin{bmatrix} 1 & 0.82 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0.82 & 1 & 0.45 & 0 & & & & \vdots \\ 0 & 0.45 & 1 & 0.82 & \ddots & & & \\ \vdots & 0 & 0.82 & 1 & 0.64 & \ddots & & \vdots \\ & & \ddots & 0.64 & 1 & 0.64 & 0 & 0 \\ & & & \ddots & 0.64 & 1 & 0.64 & 0 \\ \vdots & & & & 0 & 0.64 & 1 & 0.45 \\ 0 & \cdots & & & \cdots & 0 & 0.45 & 1 \end{bmatrix}$$

b) First we calculate $\deg(1) = 1 + 0.82 = 1.82$ and $\deg(2) = 1 + 0.82 + 0.45 = 2.27$. Then we get,

$$\mathbf{M}[1,:] = \frac{1}{1.82}[1,\ 0.82,\ 0,\ \ldots,\ 0] = [0.55,\ 0.45,\ 0,\ \ldots,\ 0]$$
$$\mathbf{M}[2,:] = \frac{1}{2.27}[0.82,\ 1,\ 0.45,\ 0,\ \ldots,\ 0] = [0.36,\ 0.44,\ 0.2,\ 0,\ \ldots,\ 0]$$

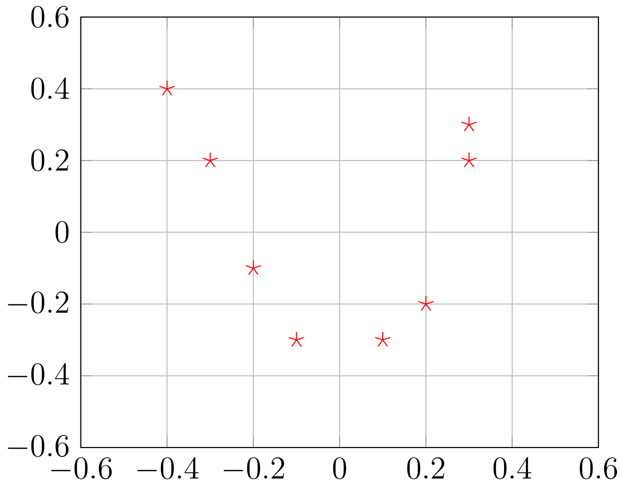
c) For $t = 0$ and $d = 2$ the diffusion map expression simplifies to

$$\phi_t^{(d)}(\mathbf{x}_i) = \begin{bmatrix} \lambda_2^t \phi_{2,i} \\ \vdots \\ \lambda_{d+1}^t \phi_{d+1,i} \end{bmatrix} = \begin{bmatrix} \phi_{2,i} \\ \vdots \\ \phi_{d+1,i} \end{bmatrix} = \begin{bmatrix} \phi_{2,i} \\ \phi_{3,i} \end{bmatrix}, i = 1, \ldots, 8\,.$$

Therefore we get the 2-D diffusion maps

$$\begin{bmatrix} \phi_t^{(d)}(\mathbf{x}_1) & \cdots & \phi_t^{(d)}(\mathbf{x}_8) \end{bmatrix} = \frac{1}{10}\begin{bmatrix} 3 & 3 & 2 & 1 & -1 & -2 & -3 & -4 \\ 3 & 2 & -2 & -3 & -3 & -1 & 2 & 4 \end{bmatrix}$$

with their corresponding 2D difussion map



d) When $t \to \infty$ all points are mapped to the zero vector.

Problem 4. (15 points)
Discriminant Analysis, (13P):

a) For two classes, the discriminant rule is

$$\mathbf{a}^T(\mathbf{x} - \frac{1}{2}(\overline{\mathbf{x}}_1 + \overline{\mathbf{x}}_2)) \gtrless 0 \,.$$

Therefore, the separating hyperplane is given by

$$\mathbf{a}^T(\mathbf{x} - \frac{1}{2}(\overline{\mathbf{x}}_1 + \overline{\mathbf{x}}_2)) = \mathbf{a}^T\mathbf{x} - \frac{1}{2}\mathbf{a}^T(\overline{\mathbf{x}}_1 + \overline{\mathbf{x}}_2) = 0 \,.$$

Since $\overline{\mathbf{x}}_1 = \frac{1}{3}(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3) = [-\frac{2}{3}, \frac{2}{3}]^T$ and $\overline{\mathbf{x}}_2 = \mathbf{x}_4 = [-1, 1]^T$, we have that

$$\overline{\mathbf{x}}_1 + \overline{\mathbf{x}}_2 = \frac{1}{3}\begin{bmatrix} 1 \\ -1 \end{bmatrix} \,.$$

This leads to the following separating hyperplane

$$\mathbf{a}^T\mathbf{x} - \frac{1}{2}\frac{1}{\sqrt{2}}\frac{1}{3}[-11]\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \mathbf{a}^T\mathbf{x} - (-\frac{1}{3\sqrt{2}}) = 0 \,,$$

thus $b = -\frac{1}{3\sqrt{2}}$. Finally, the obtained hyperplane is orthogonal to $\mathbf{a}$ and is shifted by $-\frac{1}{3\sqrt{2}}$ towards the direction of $\mathbf{a}$, then the following drawing is obtained

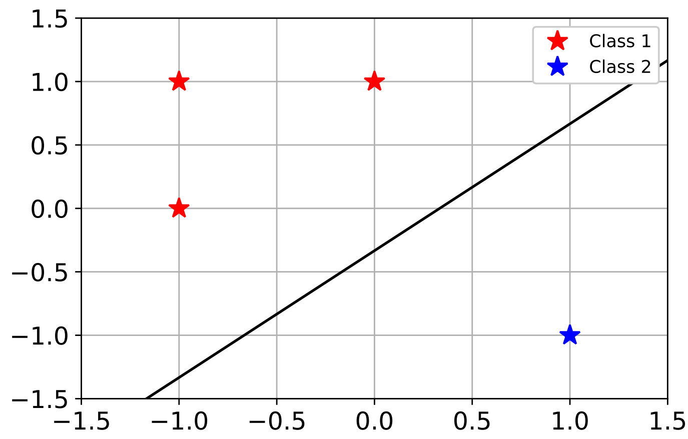


Figure 1: Separating Hyperplane

b) By definition $y_i = \mathbf{a}^T\mathbf{x}_i$, yielding

$$y_1 = \frac{1}{\sqrt{2}}[-1, 1]\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \frac{2}{\sqrt{2}}$$
$$y_2 = \frac{1}{\sqrt{2}}[-1, 1]\begin{bmatrix} -1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}}$$
$$y_3 = \frac{1}{\sqrt{2}}[-1, 1]\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}}$$
$$y_4 = \frac{1}{\sqrt{2}}[-1, 1]\begin{bmatrix} 1 \\ -1 \end{bmatrix} = -\frac{2}{\sqrt{2}} \,.$$

Then the general discriminant average is $\overline{y} = \frac{1}{4}(y_1+y_2+y_3+y_4) = \frac{1}{4\sqrt{2}}(2+1+1-2) = \frac{1}{2\sqrt{2}}$. In a simillar manner the between group averages are

$$\overline{y}_1 = \frac{1}{3}(y_1 + y_2 + y_3) = \frac{4}{3\sqrt{2}}, \quad \text{and} \quad \overline{y}_2 = y_4 = -\frac{2}{\sqrt{2}}.$$

Lets denote the sum of of squares between groups as $\gamma_B \in \mathbb{R}$. By definition we get

$$\gamma_B = \sum_{l=1}^{2} n_l(\overline{y}_l - \overline{y})^2 = 3(\frac{4}{3\sqrt{2}} - \frac{1}{2\sqrt{2}})^2 + 1(-\frac{2}{\sqrt{2}} - \frac{1}{2\sqrt{2}})^2 = \frac{25}{6}.$$

Another method to reach this solution would be to calculate

$$\mathbf{B} = \sum_{l=1}^{g} n_l(\overline{\mathbf{x}}_l - \overline{\mathbf{x}})(\overline{\mathbf{x}}_l - \overline{\mathbf{x}})^T = \cdots = \frac{25}{12}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

and then it follows

$$\gamma_B = \mathbf{a}^T\mathbf{B}\mathbf{a} = \frac{25}{6}.$$

c) Lets denote the sum of of squares within groups as $\gamma_W \in \mathbb{R}$. By definition we get

$$\begin{aligned}\gamma_W &= \sum_{l=1}^{2}\sum_{j\in C_l}(y_j - \overline{y}_l)^2 = (y_1 - \overline{y}_1)^2 + (y_2 - \overline{y}_1)^2 + (y_3 - \overline{y}_1)^2 + (y_4 - \overline{y}_2)^2 \\ &= (\frac{2}{\sqrt{2}} - \frac{4}{3\sqrt{2}})^2 + (\frac{1}{\sqrt{2}} - \frac{4}{3\sqrt{2}})^2 + (\frac{1}{\sqrt{2}} - \frac{4}{3\sqrt{2}})^2 + (-\frac{2}{\sqrt{2}} + \frac{2}{\sqrt{2}})^2 = \frac{1}{3}.\end{aligned}$$

Another method to reach this solution would be to calculate

$$\mathbf{W} = \sum_{l=1}^{g} \mathbf{X}_l^T\mathbf{E}_l\mathbf{X}_l = \cdots = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

and then it follows

$$\gamma_W = \mathbf{a}^T\mathbf{B}\mathbf{a} = \frac{1}{3}.$$

d) It is clear for Figure 1 that, for any $\epsilon > 0$, a noise vector $\eta = \mathbf{a}$ brings $\tilde{\mathbf{x}}_4$ closer to crossing the margin than any other η. Therefore, the minimum ϵ that brings $\tilde{\mathbf{x}}_4$ to the margin is

$$\begin{aligned}\mathbf{a}^T\tilde{\mathbf{x}}_4 - b &= 0 \\ \mathbf{a}^T\mathbf{x}_4 + \epsilon\mathbf{a}^T\eta - b &= 0 \\ \mathbf{a}^T\mathbf{x}_4 + \epsilon\mathbf{a}^T\mathbf{a} - b &= 0 \\ y_4 + \epsilon - b &= 0 \Rightarrow \epsilon = -(y_4 - b).\end{aligned}$$

By replacing the values of y_4 and b we obtain

$$\epsilon = -(y_4 - b) = -(-\frac{2}{\sqrt{2}} + \frac{1}{2\sqrt{2}}) = \frac{3}{2\sqrt{2}}.$$

Therefore, ϵ should be at least $\epsilon > \frac{3}{2\sqrt{2}}$ for $\tilde{\mathbf{x}}_4$ to be allocated to C_1.

Problem 5. (15 points)
Support Vector Machines: A training dataset is composed of six vectors $\mathbf{x}_i$ in two-dimensional space, $i = 1, \dots, 6$, 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:

$$\max_{\boldsymbol{\lambda}} \quad \sum_{i=1}^{6} \lambda_i - \frac{1}{2} \sum_{i=1}^{6} \sum_{j=1}^{6} y_i y_j \lambda_i \lambda_j K(\mathbf{x}_i, \mathbf{x}_j)$$

$$\text{s.t.} \quad 0 \le \lambda_i \le 2 \qquad \text{and} \qquad \sum_{i=1}^{6} \lambda_i y_i = 0.$$

The kernel function is given by:

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

The value of γ is chosen as 0.1.

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) The support vectors are all vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_5, \mathbf{x}_6$.

b) First of all, we have:

$$\mathbf{a}^\star = \sum_{i=1}^{6} \lambda_i y_i \phi(\mathbf{x}_i) = -2\phi(\mathbf{x}_1) - 0.7\phi(\mathbf{x}_2) - 1.8\phi(\mathbf{x}_3) + 2\phi(\mathbf{x}_4) + 0.5\phi(\mathbf{x}_5) + 2\phi(\mathbf{x}_6).$$

$b^\star$ can be found only for those support vectors with $\lambda \neq C = 2$.

$$\begin{aligned} b^\star &= y_3 - \sum_{i=1}^{6} \lambda_i y_i K(\mathbf{x}_i, \mathbf{x}_1) \\ &= -1 + 2K(\mathbf{x}_1, \mathbf{x}_3) + 0.74K(\mathbf{x}_2, \mathbf{x}_3) + 1.76K(\mathbf{x}_3, \mathbf{x}_3) - 2K(\mathbf{x}_4, \mathbf{x}_3) - 0.5K(\mathbf{x}_5, \mathbf{x}_3) - 2K(\mathbf{x}_6, \mathbf{x}_3) \\ &= -1 + 2 \times 0.008 + 0.74 \times 0.549 + 1.76 \times 1 - 2 \times 0.549 - 0.5 \times 0.05 - 2 \times 0 \approx 0.059. \end{aligned}$$

The exact value is $b^\star \approx 0.057928$. One can use similarly $\mathbf{x}_2$ and $\mathbf{x}_5$. Hence the classifier is given by

$$\sum_{i=1}^{6} \lambda_i y_i K(\mathbf{x}_i, \mathbf{x}) + 0.059 \gtrless 0.$$

(6P)

c) If γ is very large using the approximation we have

$$\max_{\boldsymbol{\lambda}} \quad \sum_{i=1}^{6} \lambda_i - \frac{1}{2}\sum_{i=1}^{6} y_i y_i \lambda_i^2 = -\sum_{i=1}^{6} \frac{1}{2}(\lambda_i - 1)^2$$
$$\text{s.t.} \quad 0 \leq \lambda_i \leq 2 \qquad \text{and} \qquad \sum_{i=1}^{6} \lambda_i y_i = 0.$$

The maximum in this case will be zero and is attained with $\lambda_i = 1$; note that this choice satisfies all the constraints since half of the dataset is labeled with $y_i = 1$ and the other half with $y_i = -1$ and therefore

$$\sum_{i=1}^{6} \lambda_i y_i = \sum_{i=1}^{6} y_i = 0$$

Therefore the support vector machine classifier is given by:

$$\mathbf{a}^{\star} = \sum_{i=1}^{6} y_i \phi(\mathbf{x}_i) \quad \text{and} \quad b^{\star} = 0.$$

Hence the classifier is given by

$$\sum_{i=1}^{6} y_i K(\mathbf{x}_i, \mathbf{x}) \gtrless 0.$$

However the classifier gives the output zero for each vector outside the dataset and correctly classifies the vectors inside the dataset. The support vectors include all vectors in the dataset. (3P)

Problem 6. (13 points)
Kernels for SVM:

a) (6P) See the answers below:

a) $K(\mathbf{x}_i, \mathbf{x}_j) = 1$ for all $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^p$: this is a valid kernel. The feature map is given by $\phi(\mathbf{x}) = 1$.

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 $\mathbf{x}_j = (x_j(1), \ldots, x_j(p))^T$:
This is not a valid Kernel since the kernel should be symmetric:

$$K(\mathbf{x}_1, \mathbf{x}_2) = \phi(\mathbf{x}_1)^T \phi(\mathbf{x}_2) = \phi(\mathbf{x}_2)^T \phi(\mathbf{x}_1) = K(\mathbf{x}_2, \mathbf{x}_1).$$

However this function is not symmetric.

c) $K(\mathbf{x}_i, \mathbf{x}_j) = |\|\mathbf{x}_i\|^2 - \|\mathbf{x}_j\|^2|$ for all $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^p$:
This is not a valid Kernel; If the kernel $K : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ is a valid kernel, there exists a feature function $\phi(.)$ such that

$$K(\mathbf{x}_i, \mathbf{x}_j)) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle.$$

But $K(\mathbf{x}, \mathbf{x}) = 0$ for every $\mathbf{x} \in \mathbb{R}^p$. Therefore :

$$\forall \mathbf{x} := \langle \phi(\mathbf{x}), \phi(\mathbf{x}) \rangle = 0 \implies \phi(\mathbf{x}) = 0.$$

This implies that $K(\mathbf{x}_1, \mathbf{x}_2) = 0$ for every pair of vectors $\mathbf{x}_1, \mathbf{x}_2$ which is a contradiction. One can also construct an easy example where the following matrix is not non-negative definite:

$$\mathbf{K} = \begin{pmatrix} K(\mathbf{x}_1, \mathbf{x}_1) & K(\mathbf{x}_1, \mathbf{x}_2) & \ldots & K(\mathbf{x}_1, \mathbf{x}_n) \\ K(\mathbf{x}_2, \mathbf{x}_1) & K(\mathbf{x}_2, \mathbf{x}_2) & \ldots & K(\mathbf{x}_2, \mathbf{x}_n) \\ \vdots & \vdots & \ddots & \vdots \\ K(\mathbf{x}_n, \mathbf{x}_1) & K(\mathbf{x}_n, \mathbf{x}_2) & \ldots & K(\mathbf{x}_n, \mathbf{x}_n) \end{pmatrix}.$$

b) (6P) If the kernel is given by$K(\mathbf{x}, \mathbf{y}) = 4(\mathbf{x}^T\mathbf{y})^2 + 3(\mathbf{x}^T\mathbf{y}) + 1$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$, we have:

$$\begin{aligned} K(\mathbf{x}, \mathbf{y}) &= 4(\mathbf{x}^T\mathbf{y})^2 + 3(\mathbf{x}^T\mathbf{y}) + 1 = 4(\sum_{i=1}^p x_i y_i)^2 + 3\sum_{i=1}^p x_i y_i + 1 \\ &= \sum_{i=1}^p (\sqrt{2}x_i)^2(\sqrt{2}y_i)^2 + \sum_{1 \le i < j \le p} (2\sqrt{2}x_i x_j)(2\sqrt{2}y_i y_j) + \sum_{i=1}^p (\sqrt{3}x_i)(\sqrt{3}y_i) + 1. \end{aligned}$$

The feature mapping function $\phi(\mathbf{x})$ can be written as:

$$\phi(\mathbf{x}) = (2x_1^2, \ldots, 2x_p^2, 2\sqrt{2}x_1x_2, 2\sqrt{2}x_1x_3, \ldots, 2\sqrt{2}x_{p-1}x_p, \sqrt{3}x_1, \ldots, \sqrt{3}x_p, 1).$$

The dimension is given by $p + \binom{p}{2} + p + 1 = \frac{(p+1)(p+2)}{2}$

Problem 7. (12 points)

Clustering:
Part I:

Each calculated distance $d(\mathbf{x}, \mathbf{y})$ is equivalent to 0.5P. The centers update each 2P.

a) The center of cluster 1 is $\mathbf{c}_1 = \mathbf{x}_1$, and the center of cluster 2 is $\mathbf{c}_2 = \mathbf{x}_5$.

$$d(\mathbf{c}_1, \mathbf{x}_2) = \sqrt{(3-1)^2 + (5-4)^2} = \sqrt{5} = 2.23$$
$$d(\mathbf{c}_2, \mathbf{x}_2) = \sqrt{(1+2)^2 + (4+2)^2} = \sqrt{45} = 6.7$$

$\mathbf{x}_2$ belongs to cluster 1.

$$d(\mathbf{c}_1, \mathbf{x}_3) = \sqrt{(3)^2 + (5)^2} = \sqrt{34} = 5.83$$
$$d(\mathbf{c}_2, \mathbf{x}_3) = \sqrt{(2)^2 + (2)^2} = \sqrt{8} = 2.82.$$

$\mathbf{x}_3$ belongs to cluster 2.

$$d(\mathbf{c}_1, \mathbf{x}_4) = \sqrt{(3-1)^2 + (5+1)^2} = \sqrt{40} = 6.32$$
$$d(\mathbf{c}_2, \mathbf{x}_4) = \sqrt{(1+2)^2 + (-1+2)^2} = \sqrt{10} = 3.16.$$

$\mathbf{x}_4$ belongs to cluster 2.

$$d(\mathbf{c}_1, \mathbf{x}_6) = \sqrt{(3+1)^2 + (5+4)^2} = \sqrt{97} = 9.85$$
$$d(\mathbf{c}_2, \mathbf{x}_6) = \sqrt{(-2+1)^2 + (-2+4)^2} = \sqrt{5} = 2.24.$$

$\mathbf{x}_6$ belongs to cluster 2.

b) The new center of cluster 1 is

$$\left(\tfrac{3+1}{2}, \tfrac{5+4}{2}\right) = (2, 4.5)$$

The new center of cluster 2 is

$$\left(\tfrac{0+1-2-1}{4}, \tfrac{0-1-2-4}{4}\right) = (-0.5, -1.75).$$

Part II:

<u>Iteration 1</u> (2P)
cluster C_1=$\{P_1\}$
cluster C_2=$\{P_2, P_5\}$
cluster C_3=$\{P_3\}$
cluster C_4=$\{P_4\}$

Iteration 2 (2P)

$$
\begin{aligned}
d(\mathrm{C}_1, \mathrm{C}_2) &= \frac{1}{2}(0.4 + 0.9) = 0.65 \\
d(\mathrm{C}_3, \mathrm{C}_2) &= \frac{1}{2}(0.5 + 0.2) = 0.35 \\
d(\mathrm{C}_4, \mathrm{C}_2) &= \frac{1}{2}(0.2 + 0.7) = 0.45 \\
d(\mathrm{C}_1, \mathrm{C}_4) &= 0.3
\end{aligned}
$$

cluster C_1=$\{\mathrm{P}_1, \mathrm{P}_4\}$
cluster C_2=$\{\mathrm{P}_2, \mathrm{P}_5\}$
cluster C_3=$\{\mathrm{P}_3\}$

Iteration 3 (2P)

$$
\begin{aligned}
d(\mathrm{C}_1, \mathrm{C}_2) &= \frac{1}{4}(0.9 + 0.4 + 0.2 + 0.7) = 0.55 \\
d(\mathrm{C}_2, \mathrm{C}_3) &= \frac{1}{2}(0.8 + 0.6) = 0.7 \\
d(\mathrm{C}_1, \mathrm{C}_3) &= \frac{1}{2}(0.5 + 0.2) = 0.35
\end{aligned}
$$

cluster $\mathrm{C}_2 = \{\mathrm{P}_2, \mathrm{P}_3, \mathrm{P}_5\}$
cluster $\mathrm{C}_1 = \{\mathrm{P}_1, \mathrm{P}_4\}$

Problem 8. (6 points)

Regression:

a) (4P)

$$\begin{aligned}
h_\nu(x) &= \frac{1}{1+\mathrm{e}^{0.05-0.08\times 10}} \text{ (2P)} \\
&= 0.6792 \\
\mathrm{P}(y=0 \mid x,\nu) &= 1 - h_\nu(x) \text{ (2P)} \\
&= 1 - 0.6792 \\
&= 0.3208
\end{aligned}$$

b) (2P)

$$\begin{aligned}
h_\nu(x) &= \log_{10}(1+\mathrm{e}^{-0.05+0.08\times 10}) \text{ (1P)} \\
\mathrm{P}(y=0 \mid x,\nu) &= 1 - h_\nu(x) \text{ (1P)} \\
&= 1 - \log_{10}(1+\mathrm{e}^{-0.05+0.08\times 10}) \\
&= 0.5063
\end{aligned}$$

Additional sheet

Problem:

Additional sheet

Problem:

Additional sheet

Problem: