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

1	2	3	4	$\sum$
25	25	25	25	100

Written Examination

Fundamentals of Big Data Analytics

Tuesday, March 21, 2017, 01:00 p.m.

Name: ______________________ Matr.-No.: ______________

Field of study: ______________________________________

Please pay attention to the following:

1) The exam consists of **4 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 Tuesday, the XX.XX.XX, 16:00h, on the homepage of the institute.

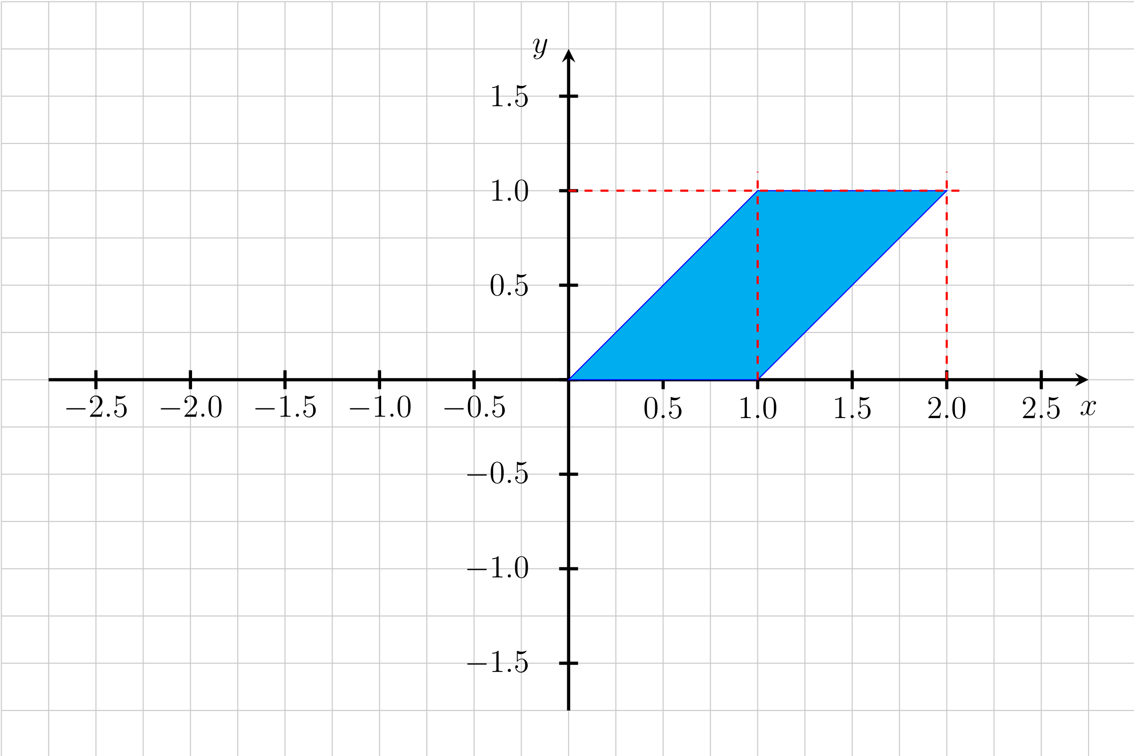
The corrected exams can be inspected on Tuesday, XX.XX.XX, 10:00h. at the seminar room 333 of the Chair for Theoretical Information Technology, Kopernikusstr. 16.

Acknowledged: ______________________________

(Signature)

Problem 1. (25 points)
Analysis of Multivariate Densities and Maximum Likelihood Estimation, (26P):

a) The support of $f_{\mathbf{Z}}(x, y)$ is given by $0 \leq y \leq 1$ and $y \leq x \leq y + 1$ for $a = 1$. This leads to the following sketch.



b) With $1 = \iint_{\mathrm{supp}\{f_{\mathbf{Z}}\}} f_{\mathbf{Z}}(x, y)\, \mathrm{d}x\, \mathrm{d}y$ we obtain

$$1 = \int_0^a \int_y^{y+a} c\, \mathrm{d}x\, \mathrm{d}y = c \int_0^a a\, \mathrm{d}y = c\, a^2 \quad \Rightarrow \quad c = \frac{1}{a^2}\,.$$

c) With $\mathrm{E}(X) = \iint_{\mathrm{supp}\{f_{\mathbf{Z}}\}} x\, f_{\mathbf{Z}}(x, y)\, \mathrm{d}x\, \mathrm{d}y$ we deduce

$$\mu_X = \int_0^1 \int_y^{y+1} x\, \mathrm{d}x\, \mathrm{d}y = \tfrac{1}{2} \int_0^1 [(y+1)^2 - y^2]\, \mathrm{d}y = \tfrac{1}{2} \int_0^1 [2y + 1]\, \mathrm{d}y = 1 \quad \Rightarrow \quad \mu_X = 1\,.$$

$$\Rightarrow \quad \mathrm{E}(\mathbf{Z}) = \tfrac{1}{2} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

d) With $\mathrm{Cov}(X, Y) = \iint_{\mathrm{supp}\{f_{\mathbf{Z}}\}} (x - \mu_X)\, (y - \mu_Y)\, f_{\mathbf{Z}}(x, y)\, \mathrm{d}x\, \mathrm{d}y$ we calculate

$$\mathrm{Cov}(Y, Y) = \int_0^1 \int_y^{y+1} \left(y - \tfrac{1}{2}\right)^2 \mathrm{d}x\, \mathrm{d}y = \int_0^1 \left(y - \tfrac{1}{2}\right)^2 \mathrm{d}y = \tfrac{1}{3} \left[\left(y - \tfrac{1}{2}\right)^3\right]_0^1 = \frac{1}{12} \quad \Rightarrow \quad \sigma_{2,2} = \tfrac{1}{12}$$

and

$$\begin{aligned} \mathrm{Cov}(X, Y) &= \int_0^1 \int_y^{y+1} (x - 1)\, \left(y - \tfrac{1}{2}\right) \mathrm{d}x\, \mathrm{d}y = \tfrac{1}{2} \int_0^1 \left[y^2 - (y-1)^2\right] \left(y - \tfrac{1}{2}\right) \mathrm{d}y \\ &= \int_0^1 \left(y - \tfrac{1}{2}\right)^2 \mathrm{d}y = \sigma_{2,2} = \frac{1}{12} \quad \Rightarrow \quad \sigma_{1,2} = \sigma_{2,1} = \tfrac{1}{12}\,. \end{aligned}$$

$$\Rightarrow \quad \mathbf{\Sigma}_{\mathbf{Z}} = \tfrac{1}{12} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

e) With $f_X(x) = \int_{\mathrm{supp}\{f_\mathbf{Z}\}} f_\mathbf{Z}(x,y)\,\mathrm{d}y$ and $f_Y(y) = \int_{\mathrm{supp}\{f_\mathbf{Z}\}} f_\mathbf{Z}(x,y)\,\mathrm{d}x$ for the marginal densities, it follows

$$\begin{aligned} 0 \le x \le 1: \quad &\Rightarrow \quad f_X(x) = \int_0^x \mathrm{d}y = x\,, \\ 1 \le x \le 2: \quad &\Rightarrow \quad f_X(x) = \int_{x-1}^1 \mathrm{d}y = 2 - x\,, \\ x \in \mathbb{R} \setminus [0,2]: \quad &\Rightarrow \quad f_X(x) = 0\,. \end{aligned}$$

and

$$\begin{aligned} 0 \le y \le 1: \quad &\Rightarrow \quad f_Y(y) = \int_y^{y+1} \mathrm{d}x = 1\,, \\ y \in \mathbb{R} \setminus [0,1]: \quad &\Rightarrow \quad f_Y(y) = 0\,. \end{aligned}$$

f) X and Y are dependent since $f_\mathbf{Z}(x,y) \neq f_X(x) f_Y(y)$.

g) Putting the substitutes $X = a\tilde{X}$ and $Y = a\tilde{Y}$ into the support defined by $0 \le y \le a$ and $y \le x \le y + a$, and dividing by a leads to a normalized support defined by $0 \le \tilde{y} \le 1$ and $\tilde{y} \le \tilde{x} \le \tilde{y} + 1$ which is the support for the case $a = 1$. Hence, in the general case $a > 0$, the random vector $\mathbf{Z}$ is just scaled by a in comparison to the case $a = 1$. This leads to

$$\Rightarrow \quad \mathrm{E}(\mathbf{Z}) = \tfrac{a}{2}\begin{pmatrix} 2 \\ 1 \end{pmatrix} \qquad \text{and} \qquad \mathbf{\Sigma}_\mathbf{Z} = \tfrac{a^2}{12}\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

for the general case.

h) Since the random vectors $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n$ are IID, the joint density is the multiplication of all single densities. So the likelihood function is given by

$$L(a, \mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n) = \prod_{i=1}^n f_\mathbf{Z}(x_i, y_i) = \frac{1}{a^{2n}}$$

i) With the log-likelihood function $\ell(a, \mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n) = \log\Big(L(a, \mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n)\Big) = -2n\log(a)$, we observe that the derivative $-2n/a$ is negative, such that ℓ is strictly decreasing in a.

j) Since ℓ is strictly decreasing in a, a smaller a leads to a larger ℓ. Due to the support of ℓ given by $0 \le y_i \le a$ and $y_i \le x_i \le y_i + a$ for all i, the parameter a is lower bounded by y_i and $x_i - y_i$. This leads to the maximum likelihood estimator $\hat{a} = \max\limits_{1 \le i \le n}\{y_i, x_i - y_i\}$.

Problem 2. (25 points)
Principal Component Analysis, (25P):

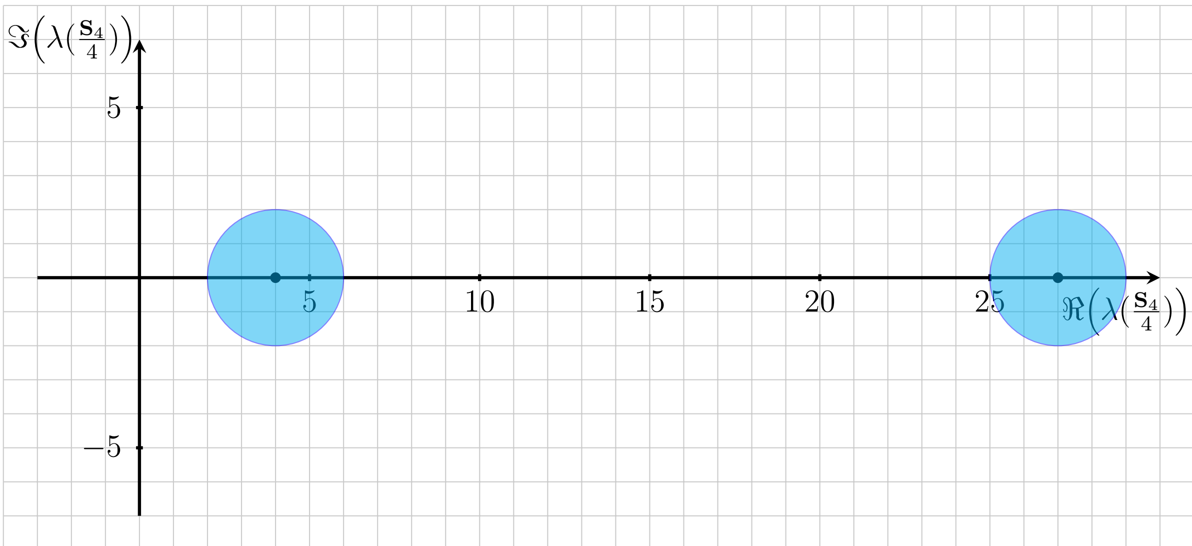
a) With $\bar{\mathbf{x}}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$, we obtain

$$\bar{\mathbf{x}}_4 = \frac{\sqrt{3}}{4}\begin{pmatrix}4+4+0+0\\4-4+8-4\end{pmatrix} = \sqrt{3}\begin{pmatrix}2\\1\end{pmatrix}.$$

b) With $\mathbf{S}_n = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^{\mathrm{T}}$, we obtain

$$\begin{aligned}\mathbf{S}_4 &= \begin{pmatrix}2\\3\end{pmatrix}\begin{pmatrix}2 & 3\end{pmatrix} + \begin{pmatrix}2\\-5\end{pmatrix}\begin{pmatrix}2 & -5\end{pmatrix} + \begin{pmatrix}-2\\7\end{pmatrix}\begin{pmatrix}-2 & 7\end{pmatrix} + \begin{pmatrix}-2\\-5\end{pmatrix}\begin{pmatrix}-2 & -5\end{pmatrix}\\ &= \begin{pmatrix}4 & 6\\6 & 9\end{pmatrix} + \begin{pmatrix}4 & -10\\-10 & 25\end{pmatrix} + \begin{pmatrix}4 & -14\\-14 & 49\end{pmatrix} + \begin{pmatrix}4 & 10\\10 & 25\end{pmatrix}\\ &= \begin{pmatrix}16 & -8\\-8 & 108\end{pmatrix} = 4\begin{pmatrix}4 & -2\\-2 & 27\end{pmatrix}.\end{aligned}$$

c) Since the normalized matrix $\frac{1}{4}\mathbf{S}_4$ is symmetric, there are only two Gerschgorin's circles. Both circles have the same radius 2 and are located at 4 and 27 on the real line.



d) The matrix $\mathbf{S}_4$ is positive definite, since both Gerschgorin's circles are strictly located on the righthand side of the complex domain.

e) The eigenvalues of $\mathbf{S}_{10k}$ are solutions of $\det(\mathbf{S}_{10k} - \mathbf{I}\lambda) = 0$. This leads to

$$\begin{vmatrix}14-\lambda & -14\\-14 & 110-\lambda\end{vmatrix} = (14-\lambda)(110-\lambda)-14^2 = \lambda^2-124\lambda+1344 = (112-\lambda)(12-\lambda) = 0\,.$$

Hence, the diagonal matrix is determined by

$$\mathbf{\Lambda} = \begin{pmatrix}112 & 0\\0 & 12\end{pmatrix}.$$

The eigenvectors $\mathbf{S}_{10k}$ are solutions of $\mathbf{S}_{10k}\mathbf{v} = \mathbf{v}\lambda$. In addition the eigenvectors should be normalized, i.e., $\|\mathbf{v}\| = 1$. We obtain

$$\begin{pmatrix}14 & -14\\-14 & 110\end{pmatrix}\begin{pmatrix}v_1\\v_2\end{pmatrix} = 112\begin{pmatrix}v_1\\v_2\end{pmatrix} \quad\Rightarrow\quad v_2 = -7v_1 \quad\text{and}\quad v_1^2+v_2^2 = 1\,,$$

which yields the normalized eigenvector $\left(\frac{1}{\sqrt{50}} \quad \frac{-7}{\sqrt{50}}\right)^{\mathrm{T}}$ for the eigenvalue 112. For the next eigenvector we only need to swap the entries of the first eigen vector and change the sign of one entry. This leads to the eigenvector $\left(\frac{7}{\sqrt{50}} \quad \frac{1}{\sqrt{50}}\right)^{\mathrm{T}}$ for the eigenvalue 12. Putting the eigenvectors together we deduce the matrix

$$\mathbf{V} = \frac{1}{\sqrt{50}} \begin{pmatrix} 1 & 7 \\ -7 & 1 \end{pmatrix} .$$

f) 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 two-dimensional sample to a one-dimensional data (k=1), we obtain

$$\mathbf{Q} = \frac{1}{\sqrt{50}} \begin{pmatrix} 1 \\ -7 \end{pmatrix} \frac{1}{\sqrt{50}} \begin{pmatrix} 1 & -7 \end{pmatrix} = \frac{1}{50} \begin{pmatrix} 1 & -7 \\ -7 & 49 \end{pmatrix} .$$

The image of $\mathbf{x}_1$ is obtained by $\frac{1}{\sqrt{50}} \begin{pmatrix} 1 & -7 \end{pmatrix} \mathbf{x}_1$ which gives $\hat{\mathbf{x}}_1 = -24\sqrt{\frac{3}{50}}$.

g) 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 $\sum_{i=1}^{k} \lambda(\mathbf{S}_n)$ of dominant eigenvalues, that is equal to 112 in the present case.

Problem 3. (25 points)
A dataset, consisting of three-dimensional vectors and their respective labels, is given below.

Data	Label	Data	Label	Data	Label
$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$	$y_1 = 1$	$\mathbf{x}_4 = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}$	$y_4 = 2$	$\mathbf{x}_7 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$	$y_7 = 3$
$\mathbf{x}_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$	$y_2 = 1$	$\mathbf{x}_5 = \begin{pmatrix} 0 \\ -2 \\ -1 \end{pmatrix}$	$y_5 = 2$	$\mathbf{x}_8 = \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}$	$y_8 = 3$
$\mathbf{x}_3 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$	$y_3 = 1$	$\mathbf{x}_6 = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$	$y_6 = 2$	$\mathbf{x}_9 = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$	$y_9 = 3$

Some general values we will need:

$$\overline{\mathbf{x}}_1 = \frac{\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3}{3} = \begin{pmatrix} 1 \\ \frac{4}{3} \\ 0 \end{pmatrix}.$$

$$\overline{\mathbf{x}}_2 = \frac{\mathbf{x}_4 + \mathbf{x}_5 \mathbf{x}_6}{3} = \begin{pmatrix} 0 \\ -\frac{4}{3} \\ -1 \end{pmatrix}.$$

$$\overline{\mathbf{x}}_3 = \frac{\mathbf{x}_7 + \mathbf{x}_8 + \mathbf{x}_9}{3} = \begin{pmatrix} -1 \\ 0 \\ \frac{4}{3} \end{pmatrix}.$$

$$\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2 = \begin{pmatrix} 1 \\ \frac{8}{3} \\ 1 \end{pmatrix}, \overline{\mathbf{x}}_2 - \overline{\mathbf{x}}_3 = \begin{pmatrix} 1 \\ -\frac{4}{3} \\ -\frac{7}{3} \end{pmatrix}.$$

a) (5P) To find $\mathbf{W}$, we have:

$$\mathbf{x_1} - \overline{\mathbf{x}}_\mathbf{1} = \begin{pmatrix} 0 \\ -\frac{1}{3} \\ 1 \end{pmatrix}, \mathbf{x_2} - \overline{\mathbf{x}}_\mathbf{1} = \begin{pmatrix} 0 \\ \frac{2}{3} \\ 0 \end{pmatrix}, \mathbf{x_3} - \overline{\mathbf{x}}_\mathbf{1} = \begin{pmatrix} 0 \\ -\frac{1}{3} \\ -1 \end{pmatrix}$$

$$\mathbf{x_4} - \overline{\mathbf{x}}_\mathbf{2} = \begin{pmatrix} 1 \\ \frac{1}{3} \\ 0 \end{pmatrix}, \mathbf{x_5} - \overline{\mathbf{x}}_\mathbf{2} = \begin{pmatrix} 0 \\ -\frac{2}{3} \\ 0 \end{pmatrix}, \mathbf{x_6} - \overline{\mathbf{x}}_\mathbf{2} = \begin{pmatrix} -1 \\ \frac{1}{3} \\ 0 \end{pmatrix}$$

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \mathbf{x}_3^T \end{pmatrix}$$

$$\mathbf{X}_2 = \begin{pmatrix} \mathbf{x}_4^T \\ \mathbf{x}_5^T \\ \mathbf{x}_6^T \end{pmatrix}$$

$$\mathbf{X}_1^T \mathbf{E}_1 \mathbf{X}_1 = \sum_{i:y_i=1} (\mathbf{x_i} - \overline{\mathbf{x}}_\mathbf{1})(\mathbf{x_i} - \overline{\mathbf{x}}_\mathbf{1})^T = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{2}{3} & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

$$\mathbf{X}_2^T\mathbf{E}_2\mathbf{X}_2 = \sum_{i:y_i=2}(\mathbf{x_i}-\overline{\mathbf{x}}_\mathbf{2})(\mathbf{x_i}-\overline{\mathbf{x}}_\mathbf{2})^T = \begin{pmatrix} 2 & 0 & 0 \\ 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Hence:

$$\mathbf{W} = \sum_{l=1}^{2}\mathbf{X}_l^T\mathbf{E}_l\mathbf{X}_l = \begin{pmatrix} 2 & 0 & 0 \\ 0 & \frac{4}{3} & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

b) (3P) Find the Fisher's linear discriminant rule for the vectors $\mathbf{x}_i$ with labels $y_i = 1$ and $y_i = 2$. Explain each step.

- Find the matrix $\mathbf{B}$ and $\mathbf{W}$. First we find the total mean for $y_i = 1, 2$:

$$\overline{\mathbf{x}} = \frac{\sum_{i=1}^{6}\mathbf{x}_i}{6} = \begin{pmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \end{pmatrix}.$$

$$(\overline{\mathbf{x}}_\mathbf{1}-\overline{\mathbf{x}})(\overline{\mathbf{x}}_\mathbf{1}-\overline{\mathbf{x}})^T = \begin{pmatrix} \frac{1}{2} \\ \frac{4}{3} \\ \frac{1}{2} \end{pmatrix}\begin{pmatrix} \frac{1}{2} & \frac{4}{3} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & \frac{2}{3} & \frac{1}{4} \\ \frac{2}{3} & \frac{16}{9} & \frac{2}{3} \\ \frac{1}{4} & \frac{2}{3} & \frac{1}{4} \end{pmatrix}.$$

$$(\overline{\mathbf{x}}_\mathbf{2}-\overline{\mathbf{x}})(\overline{\mathbf{x}}_\mathbf{2}-\overline{\mathbf{x}})^T = \begin{pmatrix} -\frac{1}{2} \\ -\frac{4}{3} \\ -\frac{1}{2} \end{pmatrix}\begin{pmatrix} -\frac{1}{2} & -\frac{4}{3} & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & \frac{2}{3} & \frac{1}{4} \\ \frac{2}{3} & \frac{16}{9} & \frac{2}{3} \\ \frac{1}{4} & \frac{2}{3} & \frac{1}{4} \end{pmatrix}.$$

Hence:

$$\mathbf{B} = \sum_{l=1}^{2}3(\overline{\mathbf{x}}_\mathbf{l}-\overline{\mathbf{x}})(\overline{\mathbf{x}}_\mathbf{l}-\overline{\mathbf{x}})^T = \begin{pmatrix} \frac{3}{2} & 4 & \frac{3}{2} \\ 4 & \frac{32}{3} & 4 \\ \frac{3}{2} & 4 & \frac{3}{2} \end{pmatrix}.$$

- Find the eigenvector of $\mathbf{W}^{-1}\mathbf{B}$ corresponding to the largest eigenvalue: Therefore :

$$\mathbf{W}^{-1} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{3}{4} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

$$\mathbf{W}^{-1}\mathbf{B} = \begin{pmatrix} \frac{3}{4} & 2 & \frac{3}{4} \\ 3 & 8 & 3 \\ \frac{3}{4} & 2 & \frac{3}{4} \end{pmatrix}.$$

As it can be seen, the matrix $\mathbf{W}^{-1}\mathbf{B}$ has the form $\begin{pmatrix} \alpha_1\mathbf{x} \\ \alpha_2\mathbf{x} \\ \alpha_3\mathbf{x} \end{pmatrix}$ and therefore its eigenvalues are given by $(0, 0, \mathrm{tr}(\mathbf{W}^{-1}\mathbf{B}))$ and its top eigenvector is given by $\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$..

Therefore the discriminant vector $\mathbf{a}$ is given by $\begin{pmatrix} 1 \\ 4 \\ 1 \end{pmatrix}$ which is the eigenvector for $\lambda = 9.5$.

- Based on $\mathbf{a}$, allocate $\mathbf{x}$ to C_1 if $\mathbf{a}^T(\mathbf{x} - \frac{1}{2}(\overline{\mathbf{x}}_\mathbf{1} + \overline{\mathbf{x}}_\mathbf{2})) > 0$, which is:

$$\begin{pmatrix} 1 & 4 & 1 \end{pmatrix}(\mathbf{x} - \begin{pmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2}. \end{pmatrix}) > 0$$

Another way:

- For two-classes, we have:

$$\mathbf{a} = \mathbf{W}^{-1}(\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2}) = \begin{pmatrix} 0.5 \\ 2 \\ 0.5 \end{pmatrix}$$

  and the discriminant rule is given by $\mathbf{a}^T(\mathbf{x} - \frac{1}{2}(\overline{\mathbf{x}}_\mathbf{1} + \overline{\mathbf{x}}_\mathbf{2})) > 0$, which is:

$$\begin{pmatrix} 0.5 & 2 & 0.5 \end{pmatrix} (\mathbf{x} - \begin{pmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2}. \end{pmatrix}) > 0$$

c) (4P) The maximum likelihood estimation of the expected values, denoted by $\hat{\mu_1}, \hat{\mu_2}, \hat{\mu_3}$, is given by $\overline{\mathbf{x}}_\mathbf{1}, \overline{\mathbf{x}}_\mathbf{2}, \overline{\mathbf{x}}_\mathbf{3}$ calculated above.

d) (5P) The maximum likelihood estimation of the covariance matrix, $\hat{\mathbf{\Sigma}}$, is given by $\frac{\mathbf{W}}{n}$ where:

$$\mathbf{W} = \sum_{l=1}^{3} \mathbf{X}_l^T \mathbf{E}_l \mathbf{X}_l.$$

n is equal to 9 in this problem. We need only to find $\mathbf{X}_3^T \mathbf{E}_3 \mathbf{X}_3$:

$$\mathbf{X}_3^T \mathbf{E}_3 \mathbf{X}_3 = \sum_{i:y_i=3} (\mathbf{x_i} - \overline{\mathbf{x}}_\mathbf{3})(\mathbf{x_i} - \overline{\mathbf{x}}_\mathbf{3})^T = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \frac{2}{3} \end{pmatrix}$$

Then:

$$\hat{\mathbf{\Sigma}} = \frac{1}{9} \begin{pmatrix} 2 & 0 & 0 \\ 0 & \frac{10}{3} & 0 \\ 0 & 0 & \frac{8}{3} \end{pmatrix}.$$

e) (3P) To find $\hat{\mathbf{\Sigma}}$, we use its ML-estimation which is $\frac{1}{6}\mathbf{W}$ found using only those vectors with label 2 and 3.

$$\hat{\mathbf{\Sigma}} = \frac{1}{6} \begin{pmatrix} 2 & 0 & 0 \\ 0 & \frac{8}{3} & 0 \\ 0 & 0 & \frac{2}{3} \end{pmatrix}.$$

f) (5P) The ML rule allocates $\mathbf{x}$ to the class C_1 if

$$\alpha^T(\mathbf{x} - \hat{\mu}) > 0,$$

where $\alpha = \hat{\mathbf{\Sigma}}^{-1}(\hat{\mu_1} - \hat{\mu_2})$ and $\mu = \frac{1}{2}(\hat{\mu_1} + \hat{\mu_2})$. The estimations $\hat{\mu_1}$ and $\hat{\mu_2}$ are available from the previous problems. Finally:

$$\alpha = \hat{\mathbf{\Sigma}}^{-1}(\overline{\mathbf{x}_2} - \overline{\mathbf{x}_3}) = \frac{1}{6} \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{7}{2} \end{pmatrix}.$$

And then : $\alpha^T(\mathbf{x} - \frac{1}{2}(\overline{\mathbf{x}}_\mathbf{3} + \overline{\mathbf{x}}_\mathbf{2})) > 0$, which is:

$$\frac{1}{6} \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & -\frac{7}{2} \end{pmatrix} (\mathbf{x} - \begin{pmatrix} -\frac{1}{2} \\ -\frac{2}{3} \\ \frac{1}{6}. \end{pmatrix}) > 0$$

Problem 4. (25 points)
(25P)

a) (3P) b is given as $-\frac{1}{2}\mathbf{a}^T(\mathbf{x}_1+\mathbf{x}_2)=3$.

b) (4P) Supporting vectors are those with $\lambda_i \neq 0$, which are $\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5$.

$\mathbf{x}_i$	y_i	λ_i	$\mathbf{x}_i$	y_i	λ_i
$\mathbf{x}_1=\begin{pmatrix}1\\1\end{pmatrix}$	$y_1=-1$	$\lambda_1=0$	$\mathbf{x}_4=\begin{pmatrix}0.5\\-0.5\end{pmatrix}$	$y_4=1$	$\lambda_4=4.73$
$\mathbf{x}_2=\begin{pmatrix}2\\0\end{pmatrix}$	$y_2=-1$	$\lambda_2=0.67$	$\mathbf{x}_5=\begin{pmatrix}-2\\1\end{pmatrix}$	$y_5=1$	$\lambda_5=0.94$
$\mathbf{x}_3=\begin{pmatrix}0\\0\end{pmatrix}$	$y_3=-1$	$\lambda_3=5$	$\mathbf{x}_6=\begin{pmatrix}0\\-1\end{pmatrix}$	$y_6=1$	$\lambda_1=0$

c) (6P) For those vectors, the normal vector of the hyperplane is obtained as:

$$\mathbf{a}=\sum_{i=1}^{6}\lambda_i y_i \mathbf{x}_i=\lambda_2 y_2\mathbf{x}_2+\lambda_3 y_3\mathbf{x}_3+\lambda_4 y_4\mathbf{x}_4+\lambda_5 y_5\mathbf{x}_5$$

$$\mathbf{a}=-0.67\begin{pmatrix}2\\0\end{pmatrix}-5\begin{pmatrix}0\\0\end{pmatrix}+4.73\begin{pmatrix}0.5\\-0.5\end{pmatrix}+0.94\begin{pmatrix}-2\\1\end{pmatrix}=\begin{pmatrix}-0.86\\-1.43\end{pmatrix}$$

To find b, take two support vectors $\mathbf{x}_k$ and $\mathbf{x}_l$ with $y_k=1$ and $y_l=-1$ with $0<\lambda<5$. For these support vectors, we have $y_i(\mathbf{a}^T\mathbf{x}_i+b)=1$. Hence:

$$b^\star=\frac{-1}{2}\mathbf{a}^{\star T}(\mathbf{x}_k+\mathbf{x}_l)=-\frac{1}{2}\begin{pmatrix}-0.86 & -1.43\end{pmatrix}\left(\begin{pmatrix}2\\0\end{pmatrix}+\begin{pmatrix}-2\\1\end{pmatrix}\right)=\frac{1.43}{2}=0.715. \tag{1}$$

d) (6P)

Suppose that a kernel is given by $K(\mathbf{x},\mathbf{y})=(2\mathbf{x}^T\mathbf{y}+1)^2$ for $\mathbf{x},\mathbf{y}\in\mathbb{R}^p$. Write the kernel as

$$\begin{aligned}K(\mathbf{x},\mathbf{y})=(2\mathbf{x}^T\mathbf{y}+1)^2&=(2\sum_{i=1}^{p}x_iy_i+1)^2\\&=4\sum_{i=1}^{p}x_i^2y_i^2+8\sum_{1\le i<j\le p}x_ix_jy_iy_j+4\sum_{i=1}^{p}x_iy_i+1,\end{aligned}$$

therefore $\phi(\mathbf{x})$ can be written as:

$$\phi(\mathbf{x})=(2x_1^2,\ldots,2x_p^2,2x_1,\ldots,2x_p,1,\sqrt{8}x_1x_2,\sqrt{8}x_1x_3,\ldots,\sqrt{8}x_{p-1}x_p).$$

The dimension of feature space is $p+p+1+\frac{p(p-1)}{2}=\frac{(p+1)(p+2)}{2}$.

e) (3P) The Kernel calssifer replaces the inner product in dual problem:

$$\begin{aligned}\max\quad&\sum_{i=1}^{n}\lambda_i-\frac{1}{2}\sum_{i,j}y_iy_j\lambda_i\lambda_jK(\mathbf{x}_i,\mathbf{x}_j)\\\text{s.t.}\quad&0\le\lambda_i\\&\sum_{i=1}^{n}\lambda_iy_i=0.\end{aligned}$$

For the proposed K we have:

$$\begin{aligned}
\max \quad & \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i,j} y_i y_j \lambda_i \lambda_j \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \\
\text{s.t.} \quad & 0 \leq \lambda_i \\
& \sum_{i=1}^{n} \lambda_i y_i = 0.
\end{aligned}$$

f) (3P) From this optimization problem, the vector $\phi(\mathbf{a})$ in the feature space is obtained as:

$$\phi(\mathbf{a}) = \sum_{i=1}^{n} \lambda_i y_i \phi(\mathbf{x}_i).$$

and for two vectors with $y_k = 1$ and $y_l = -1$ and $0 < \lambda$.

$$\begin{aligned}
b^\star &= \frac{-1}{2} \phi(\mathbf{a}^\star)^T (\phi(\mathbf{x}_k) + \phi(\mathbf{x}_l)) \\
&= \frac{-1}{2} (\sum_{i=1}^{n} \lambda_i y_i \phi(\mathbf{x}_i))^T (\phi(\mathbf{x}_k) + \phi(\mathbf{x}_l)) \\
&= \frac{-1}{2} \sum_{i=1}^{n} \lambda_i y_i (\phi(\mathbf{x}_i)^T \phi(\mathbf{x}_k) + \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_l)) \\
&= \frac{-1}{2} \sum_{i=1}^{n} \lambda_i y_i (K(\mathbf{x}_i, \mathbf{x}_k) + K(\mathbf{x}_i, \mathbf{x}_l)).
\end{aligned}$$

The kernel classifier is given as:

$$\phi(\mathbf{a})^T \phi(\mathbf{x}) + b^\star \geq 1 \implies \mathbf{x} \in \mathcal{C}_1$$

$$\phi(\mathbf{a})^T \phi(\mathbf{x}) + b^\star \leq -1 \implies \mathbf{x} \in \mathcal{C}_2$$

where

$$\phi(\mathbf{a})^T \phi(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i y_i K(\mathbf{x}_i, \mathbf{x})$$