Data Analytics and Machine Learning
Department of Informatics
Technical University of Munich

Esolution
Place student sticker here

Note:

- During the attendance check a sticker containing a unique code will be put on this exam.
- This code contains a unique number that associates this exam with your registration number.
- This number is printed both next to the code and to the signature field in the attendance check list.

Maschinelles Lernen

Exam:	IN2064 / Endterm	**Date:**	Thursday 17th February, 2022
Examiner:	Prof. Dr. Stephan Günnemann	**Time:**	17:00 – 19:00

Working instructions

- You have to sign the code of conduct. (Typing your name is fine).
- You have to either print this document, solve the problems on paper and then scan your solutions OR paste scans/pictures of your handwritten, on-paper solutions into the solution boxes in this PDF.
- You must not use any other means of creating a submission (e.g. digital pen on a tablet).
- Make sure that the **QR codes are visible** on every uploaded page. Otherwise, we cannot grade your submission.
- **You must solve the specified version of the problem**. Different problems may have different versions: e.g. Problem 1 (Version A), Problem 5 (Version C), etc. If you solve the wrong version you get **zero** points.
- Only write on the provided sheets, **submitting your own additional sheets is not possible**.
- The last pages (after problem 11) can be used as scratch paper.
- All sheets (save for empty scratch paper) have to be submitted to the upload queue. Missing pages will be considered empty.
- **Only use a black or blue color (no red or green)! Pencils are allowed.**
- **For problems that say "Justify your answer" you only get points if you provide a valid explanation.**
- **For problems that say "Derive" you only get points if you provide a valid mathematical derivation.**
- **For problems that say "Prove" you only get points if you provide a valid mathematical proof.**
- If a problem does not say "Justify your answer", "Derive" or "Prove" it's sufficient to only provide the correct answer.
- Instructor announcements and clarifications will be posted **on Piazza** with email notifications.
- Exercise duration - 120 minutes.

Problem 1 (All versions) (10 credits)

0 1 2 3 4 5 6 7 8 9 10

The posterior distribution is proprotional to:

$$
\begin{aligned}
\mathbb{P}(\{x_1, ..., x_N\} \mid \theta) \times \mathbb{P}(\theta \mid \lambda, \alpha) &= \prod_i \mathbb{P}(x_i \mid \theta) \times \mathbb{P}(\theta \mid \lambda, \alpha) \\
&= \frac{1}{\theta^N} 1_{\max(x_i) \leq \theta} \frac{1}{\theta^{\alpha+1}} 1_{\max(\lambda) \leq \theta} \alpha \lambda^{\alpha} \\
&\propto \frac{1}{\theta^{N+\alpha+1}} 1_{\max(x_1, ..., x_N, \lambda) \leq \theta}
\end{aligned}
$$

.
.
.
.
.

We recognize that the posterior distibution is $\mathbb{P}(\theta \mid \{x_1, ..., x_N\}, \lambda, \alpha) = \mathbb{P}(\theta \mid \lambda_{\text{new}}, \alpha_{\text{new}}) = \mathit{Pareto}(\lambda_{\text{new}} = \max(x_1, ..., x_N, \lambda), \alpha_{\text{new}} = N + \alpha)$.

Problem 2 (All versions) (8 credits)

0
1
2
3
4
5
6
7
8

We write the optimal solution of the new problem:

$$\mathbf{w}_{new}^* = (\mathbf{X}_{new}^T \mathbf{X}_{new})^{-1} \mathbf{X}_{new}^T \frac{1}{\sigma} \mathbf{y}$$

We want to have:

$$(\mathbf{X}_{new}^T \mathbf{X}_{new})^{-1} \mathbf{X}_{new}^T \frac{1}{\sigma} \mathbf{y} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

This is achieved if $\mathbf{X}_{new} = \frac{1}{\sigma} \mathbf{X}$

.
.
.

Problem 3 (Version A) (3 credits)

0
1

a)

Distance	Assignment (a-c)
L_2-distance: $\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_2 = \sqrt{\sum_i \left(\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}\right)^2}$	b
L_1-distance: $\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_1 = \sum_i \left\|\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}\right\|$	a
L_∞-distance: $\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_\infty = \max_i \left\|\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}\right\|$	c

0
1

b)

Class 2 (triangle)

0
1

c)

Class 1 (x)

Problem 3 (Version B) (3 credits)

a)

Distance	Assignment (a-c)
L_2-distance: $\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_2 = \sqrt{\sum_i \left(\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}\right)^2}$	b
L_∞-distance: $\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_\infty = \max_i \left\|\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}\right\|$	c
L_1-distance: $\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_1 = \sum_i \left\|\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}\right\|$	a

0
1

b)

Class 2 (x)

0
1

c)

Class 1 (triangle)

0
1

Problem 3 (Version C) (3 credits)

0
1
a)

Distance	Assignment (a-c)
L_∞-distance: $\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_\infty = \max_i \left\|\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}\right\|$	c
L_2-distance: $\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_2 = \sqrt{\sum_i \left(\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}\right)^2}$	b
L_1-distance: $\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_1 = \sum_i \left\|\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}\right\|$	a

0
1
b)

Class 2 (triangle)

0
1
c)

Class 1 (x)

Problem 3 (Version D) (3 credits)

a)

0
1

Distance	Assignment (a-c)
L_∞-distance: $\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_\infty = \max_i \left|\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}\right|$	c
L_1-distance: $\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_1 = \sum_i \left|\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}\right|$	a
L_2-distance: $\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_2 = \sqrt{\sum_i \left(\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}\right)^2}$	b

b)

0
1

Class 2 (x)

c)

0
1

Class 1 (triangle)

Problem 4 (Version A) (6 credits)

0 1 2

a)

1. Two isotropic bivariate Gaussian distributions with identical covariance matrices
2. Inconclusive, since both models' assumptions match the data.

0 1 2

b)

1. Identical covariance matrices but correlated features
2. Linear Discriminant Analysis since the covariance matrices match for both classes and the features are correlated (alternative: Naïve Bayes' conditional independence assumption is clearly violated).

0 1 2

c)

1. Covariance matrices do not match and no apparent correlation between features
2. Naïve Bayes' since the features are conditionally independent and variances mismatch between classes $\Rightarrow$ quadratic decision boundary (alternative: Linear Discriminant Analysis' assumption of equal covariance matrices is clearly violated).

Problem 4 (Version B) (6 credits)

a)

0 1 2

1. Identical covariance matrices but correlated features
2. Linear Discriminant Analysis since the covariance matrices match for both classes and the features are correlated (alternative: Naïve Bayes' conditional independence assumption is clearly violated).

b)

0 1 2

1. Two isotropic bivariate Gaussian distributions with identical covariance matrices
2. Inconclusive, since both models' assumptions match the data.

c)

0 1 2

1. Covariance matrices do not match and no apparent correlation between features
2. Naïve Bayes' since the features are conditionally independent and variances mismatch between classes $\Rightarrow$ quadratic decision boundary (alternative: Linear Discriminant Analysis' assumption of equal covariance matrices is clearly violated).

Problem 4 (Version C) (6 credits)

0 1 2

a)

1. Covariance matrices do not match and no apparent correlation between features
2. Naïve Bayes' since the features are conditionally independent and variances mismatch between classes $\Rightarrow$ quadratic decision boundary (alternative: Linear Discriminant Analysis' assumption of equal covariance matrices is clearly violated).

0 1 2

b)

1. Identical covariance matrices but correlated features
2. Linear Discriminant Analysis since the covariance matrices match for both classes and the features are correlated (alternative: Naïve Bayes' conditional independence assumption is clearly violated).

0 1 2

c)

1. Two isotropic bivariate Gaussian distributions with identical covariance matrices
2. Inconclusive, since both models' assumptions match the data.

Problem 4 (Version D) (6 credits)

a)

0 1 2

1. Covariance matrices do not match and no apparent correlation between features
2. Naïve Bayes' since the features are conditionally independent and variances mismatch between classes $\Rightarrow$ quadratic decision boundary (alternative: Linear Discriminant Analysis' assumption of equal covariance matrices is clearly violated).

b)

0 1 2

1. Two isotropic bivariate Gaussian distributions with identical covariance matrices
2. Inconclusive, since both models' assumptions match the data.

c)

0 1 2

1. Identical covariance matrices but correlated features
2. Linear Discriminant Analysis since the covariance matrices match for both classes and the features are correlated (alternative: Naïve Bayes' conditional independence assumption is clearly violated).

Problem 5 (All versions) (10 credits)

0 1 2 3 4

a)

From the lecture we know: Let $f_1 : \mathbb{R}^d \to \mathbb{R}$ and $f_2 : \mathbb{R}^d \to \mathbb{R}$ be convex functions, and $g : \mathbb{R}^d \to \mathbb{R}$ be a concave function, then: (1) $h(\boldsymbol{x}) = f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})$ is convex, (2) $h(\boldsymbol{x}) = \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\}$ is convex, (3) $h(\boldsymbol{x}) = c \cdot f_1(\boldsymbol{x})$ is convex if $c \geq 0$, (4) $h(\boldsymbol{x}) = c \cdot g(\boldsymbol{x})$ is convex if $c \leq 0$, (5) $h(\boldsymbol{x}) = f_1(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$ is convex ($\boldsymbol{A}$ matrix, $\boldsymbol{b}$ vector), and (6) $h(\boldsymbol{x}) = m(f_1(\boldsymbol{x}))$ is convex if $m : \mathbb{R} \to \mathbb{R}$ is convex and nondecreasing.

Consider the function $e_i(\mathbf{x}) = x_i$ which is clearly convex in $\mathbf{x}$ since it is constant in all dimensions but the i-th, in which it is linear. From the definition of convexity, it is easy to see that $e_i(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda e_i(\mathbf{x}) + (1-\lambda)e_i(\mathbf{y})$ holds.
(By induction, rule (2) also holds for more than two arguments.) Thus, by rule (2) $\max_{i=1,\ldots,n} x_i$ is convex (plugging in $e_i(\mathbf{x})$ for $f_i(\mathbf{x})$).
Equivalently $\min_{i=1,\ldots,n} x_i$ is concave. By rule (4) it follows that $-\min_{i=1,\ldots,n} x_i$ is convex.
Last, by rule (1) it follows that $f(\mathbf{x})$ is convex in $\mathbf{x}$.

0 1 2 3 4 5 6

b)

Option 1: direct application of the convexity definition

For two arbitrary $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, by definition of $g(x)$: $t_x^* = \text{median}(\mathbf{x}) = \arg\min_t \|\mathbf{x} - t \cdot \mathbf{1}\|_1$ and $t_y^* = \text{median}(\mathbf{y}) = \arg\min_t \|\mathbf{y} - t \cdot \mathbf{1}\|_1$. Moreover, $t_\lambda^* = \text{median}(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) = \arg\min_t \|\lambda\mathbf{x} + (1-\lambda)\mathbf{y} - t \cdot \mathbf{1}\|_1$ for an arbitrary $\lambda \in [0, 1]$. Also note that $\sum_{i=1}^n |x_i - t| = \|\mathbf{x} - t\mathbf{1}\|_1$.

$$\begin{aligned}
\lambda g(\mathbf{x}) + (1-\lambda)g(\mathbf{y}) &= \lambda\frac{1}{n}\|\mathbf{x} - t_x^*\mathbf{1}\|_1 + (1-\lambda)\frac{1}{n}\|\mathbf{y} - t_y^*\mathbf{1}\|_1 \\
&= \frac{1}{n}\left[\|\lambda\mathbf{x} - \lambda t_x^*\mathbf{1}\|_1 + \|(1-\lambda)\mathbf{y} - (1-\lambda)t_y^*\mathbf{1}\|_1\right] \\
&\overset{(1)}{\geq} \frac{1}{n}\|\lambda\mathbf{x} - \lambda t_x^*\mathbf{1} + (1-\lambda)\mathbf{y} - (1-\lambda)t_y^*\mathbf{1}\|_1 \\
&\overset{(2)}{\geq} \frac{1}{n}\|\lambda\mathbf{x} + (1-\lambda)\mathbf{y} - t_\lambda^*\mathbf{1}\|_1 \\
&= g(\lambda\mathbf{x} + (1-\lambda)\mathbf{y})
\end{aligned}$$

(1) follows from triangle inequality ($\|\mathbf{a}\| + \|\mathbf{b}\| \geq \|\mathbf{a} + \mathbf{b}\|$) and (2) from the definition of t_λ^* above. Since $\lambda g(\mathbf{x}) + (1-\lambda)g(\mathbf{y}) \geq g(\lambda\mathbf{x} + (1-\lambda)\mathbf{y})$ it follows that $g(\mathbf{x})$ is convex.

Option 2: argument via convexity preserving operations (as much as possible)

An integral part is to analyze the convexity of $\sum_{i=1}^{n} |x_i - t| = \|\mathbf{x} - t\mathbf{1}\|_1$. First, observe that $\mathbf{x} - t\mathbf{1}$ is a linear function in $\mathbf{x}$ (as well as t). Equivalently, we can write $\mathbf{I}\mathbf{x} - t\mathbf{1} = \mathbf{A}\mathbf{x} + \mathbf{b}$. Moreover, $\|\mathbf{x}\|_1$ is convex but *not nondecreasing for* $\mathbf{x} \in \mathbb{R}^n$ (i.e. rule 6 does not apply). From rule 5 it follows that $\|\mathbf{x} - t\mathbf{1}\|_1$ is convex in $\mathbf{x}$ (as well as t).

By rule (3) $g(\mathbf{x})$ is convex if $\sum_{i=1}^{n} |x_i - \text{median}(\mathbf{x})|$ is convex since $\frac{1}{n} > 0$.

$\text{median}(\mathbf{x}) = \arg\min_{t\in\mathbb{R}} \|\mathbf{x} - t\mathbf{1}\|_1$. Hence, we can write $g(\mathbf{x}) = \frac{1}{n}\|\mathbf{x} - \text{median}(\mathbf{x})\|_1 = \frac{1}{n}\min_{t\in\mathbb{R}} \|\mathbf{x} - t\mathbf{1}\|_1$.

Now it is left to proof that $g(\mathbf{x}) = \min_{t\in\mathbb{R}} f(\mathbf{x}, t)$ is convex for a function $f(\mathbf{x}, t)$ that convex in both $\mathbf{x}$ and t . For arbitrary $\mathbf{x}_1$ and $\mathbf{x}_2$ we define $t_1 = \arg\min_{t\in\mathbb{R}} f(\mathbf{x}_1, t)$ and $t_2 = \arg\min_{t\in\mathbb{R}} f(\mathbf{x}_2, t)$. For an arbitrary $\lambda \in [0, 1]$:

$$\begin{aligned} g(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) &= \min_{t\in\mathbb{R}} f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2, t) \\ &\le f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2, \lambda t_1 + (1-\lambda)t_2) \\ &\le \lambda f(\mathbf{x}_1, t_1) + (1-\lambda) f(\mathbf{x}_2, t_2) \\ &= \lambda g(\mathbf{x}_1) + (1-\lambda) g(\mathbf{x}_2) \end{aligned}$$

Hence, $g(\mathbf{x})$ is a convex function.

Problem 6 (Version A) (8 credits)

0
1
2
3
4

a)

```
out = x * np.sin(y)
```

Also correct: `out = x * sin(y)`

0
1
2
3
4

b)

```
N = len(z)
d_z = np.ones_like(z) / N * d_out
```

Also correct: `1 / N * d_out` with `np.repeat` or similar to make vector

Problem 6 (Version B) (8 credits)

a)

```
out = np.exp(x) / np.exp(y)
```

Also correct: `out = np.exp(x - y)` and without np, or writing e^{x-y}

0 1 2 3 4

b)

```
d_z = np.ones_like(z) * d_out
```

0 1 2 3 4

Problem 6 (Version C) (8 credits)

0 1 2 3 4

a)

```
out = np.sin(x * y)
```

Also correct without numpy, writing only `sin(x * y)`

0 1 2 3 4

b)

```
d_z = np.prod(z) / z * d_out
```

Also correct: any code that computes the product $\prod_{i \neq j} z_i$

Problem 6 (Version D) (8 credits)

a)

0
1
2
3
4

```
out = x + x * y
```

b)

0
1
2
3
4

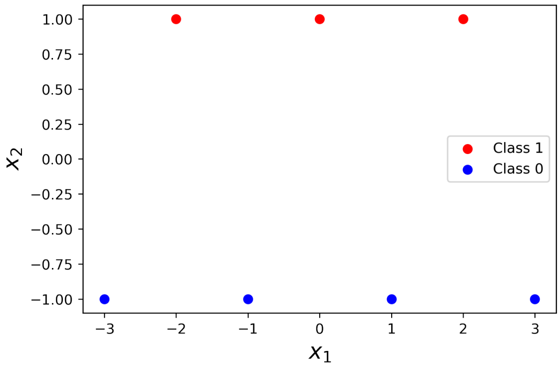
```
d_z = 2 * z * d_out
```

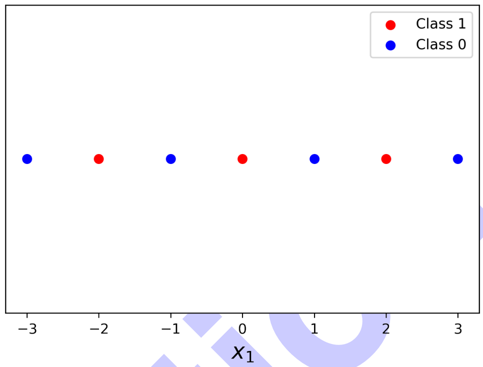
Problem 7 (All versions) (12 credits)

0 1 2 3 4 5 6

a)

Consider the 2-dimensional dataset in the left figure, which is clearly linearly separable.





The features are already uncorrelated, meaning PCA will not change the dataset. When we remove the lower-variance dimension x_2, the data is clearly no longer linearly separable, as shown in the right figure.

0 1 2 3 4 5 6

b)

Proof by contradiction.
Assume that $(\mathbf{X}, \mathbf{y})$ was not linearly separable, but $\left(\tilde{\mathbf{X}}, \mathbf{y}\right)$ was linearly separable.
Let $\Gamma \in \mathbb{R}^{D \times K}$ be the top-K eigenvectors of the covariance matrix, i.e. the matrix we use to perform dimensionality reduction.
Since $\left(\tilde{\mathbf{X}}, \mathbf{y}\right)$ is linearly separable, there is a linear classifier $\mathbb{I}\left[\mathbf{w}^T x \geq b\right]$ with parameters $\mathbf{w} \in \mathbb{R}^k$ and $b \in \mathbb{R}$ that correctly classifies all points in $\left(\tilde{\mathbf{X}}, \mathbf{y}\right)$.
Based on the definition of $\tilde{\mathbf{X}}$, i.e. $\tilde{\mathbf{X}} = \Gamma^T \mathbf{X}$, this means that any point in $(\mathbf{X})$ can be correctly classified by first performing dimensionality reduction with PCA and then applying the linear classifier specified above.
However, this procedure in itself is a linear classifier $\mathbb{I}\left[\left(\mathbf{w}^T \Gamma^T\right) x \geq b\right]$ with weight vector $\mathbf{w}' = \Gamma \cdot w$.
This contradicts our assumption that $(\mathbf{X}, \mathbf{y})$ is not linearly separable.

Problem 8 (Version A) (6 credits)

0
1
2
3
4
5
6

- Loss curve A corresponds to model 1: Both the training and test losses a high — the model *underfits*. This may happen if k is too low.
- Loss curve C corresponds to model 3: The train loss is low, but the test loss increases sharply after a few iterations — this is a clear example of *overfitting*. This may happen if k is too high.
- Loss curve B corresponds to model 2: by exclusion.

Problem 8 (Version B) (6 credits)

0
1
2
3
4
5
6

- Loss curve B corresponds to model 1: Both the training and test losses a high — the model *underfits*. This may happen if k is too low.
- Loss curve C corresponds to model 3: The train loss is low, but the test loss increases sharply after a few iterations — this is a clear example of *overfitting*. This may happen if k is too high.
- Loss curve A corresponds to model 2: by exclusion.

Problem 8 (Version C) (6 credits)

0
1
2
3
4
5
6

- Loss curve B corresponds to model 1: Both the training and test losses a high — the model *underfits*. This may happen if k is too low.
- Loss curve A corresponds to model 3: The train loss is low, but the test loss increases sharply after a few iterations — this is a clear example of *overfitting*. This may happen if k is too high.
- Loss curve C corresponds to model 2: by exclusion.

Problem 8 (Version D) (6 credits)

0
1
2
3
4
5
6

- Loss curve C corresponds to model 1: Both the training and test losses a high — the model *underfits*. This may happen if k is too low.
- Loss curve B corresponds to model 3: The train loss is low, but the test loss increases sharply after a few iterations — this is a clear example of *overfitting*. This may happen if k is too high.
- Loss curve A corresponds to model 2: by exclusion.

Problem 9(All versions) (12 credits)

First, we write down the objective function

$$
\begin{aligned}
\underset{\mathbf{Z}\sim\gamma_t(\mathbf{Z})}{\mathbb{E}}[\log p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_K)] &= \sum_{i=1}^{N}\sum_{k=1}^{K}\gamma_t(z_i = k)\log\left(\frac{1}{K}\prod_{d=1}^{D}\frac{(x_{id}\mu_{kd})^{(x_{id}-1)}\exp(-\mu_{kd}x_{id})}{x_{id}!}\right) \\
&= C + \sum_{i=1}^{N}\sum_{k=1}^{K}\gamma_t(z_i = k)\sum_{d=1}^{D}((x_{id}-1)\log\mu_{kd} - \mu_{kd}x_{id}) \\
&= C + \sum_{k=1}^{K}\sum_{d=1}^{D}\underbrace{\sum_{i=1}^{N}\gamma_t(z_i = k)((x_{id}-1)\log\mu_{kd} - \mu_{kd}x_{id})}_{=:\mathcal{L}_{kd}} \\
&= C + \sum_{k=1}^{K}\sum_{d=1}^{D}\mathcal{L}_{kd}
\end{aligned}
$$

We see that the objective can be decomposed as a sum, where each parameter μ_{kd} is only encountered in a single term $\mathcal{L}_{kd}$.

Therefore, to find the optimal μ_{kd}, we only need to maximize the respective term $\mathcal{L}_{kd}$.

For this, we find the derivative of $\mathcal{L}_{kd}$ w.r.t. μ_{kd} and set it to zero:

$$
\begin{aligned}
\frac{\partial\mathcal{L}_{kd}}{\partial\mu_{kd}} &= \frac{\partial}{\partial\mu_{kd}}\left(\sum_{i=1}^{N}\gamma_t(z_i = k)((x_{id}-1)\log\mu_{kd} - \mu_{kd}x_{id})\right) \\
&= \sum_{i=1}^{N}\gamma_t(z_i = k)\left(\frac{x_{id}-1}{\mu_{kd}} - x_{id}\right) \overset{!}{=} 0.
\end{aligned}
$$

Setting the gradient to zero and solving for μ_{kd}, we obtain the update

$$
\mu_{kd} = \frac{\sum_{i=1}^{N}\gamma_t(z_i = k)(x_{id}-1)}{\sum_{i=1}^{N}\gamma_t(z_i = k)x_{id}}.
$$

Problem 10 (Version A) (10 credits)

0
1
2
3
4
5
6

a)

We need to solve $\max_{\mathbf{x} \simeq \mathbf{x}'} ||f(\mathbf{x}) - f(\mathbf{x}')||_1$.
Using the definition of f, this can be written as

$$\max_{\mathbf{x} \simeq \mathbf{x}'} ||\mathbf{A}\left(\mathbf{x} - \mathbf{x}'\right)||_1$$

Due to the definition of $\simeq$, $\mathbf{x}$ and $\mathbf{x}'$ can only differ in one component and at most by 1. Thus, the above is equivalent to

$$\max_{j \in \{1,2,3,4\}} \max_{c \in [-1,1]} ||\mathbf{A} \cdot (c \cdot e_j)||_1 ,$$

where e_j is the unit vector with non-zero entry in dimension j.
The product with the unit vector is equivalent to summing over the j'th column of A, i.e.

$$\max_{j \in \{1,2,3,4\}} \max_{c \in [-1,1]} \left|\left| c \cdot \sum_{d=1}^{4} A_{:,d} \right|\right|_1 .$$

The function evidently has its maximum at $c = \pm 1$ and thus we have

$$\Delta_1 = \max_{j \in \{1,2,3,4\}} \left|\left| \sum_{d=1}^{4} A_{:,d} \right|\right|_1 .$$

To summarize: The maximum is achieved by changing dimension corresponding to the column whose sum has the largest absolute value. In our case, the Δ_1-sensitivity is 24.

b)

$\frac{24}{\frac{1}{2}}$

c)

This holds due to group-privacy
$\mathbf{x} \simeq \mathbf{x}'$ means that $\mathbf{x}$ and $\mathbf{x}'$ differ up to 1 in exactly one dimension. $\mathbf{x} \simeq \infty \mathbf{x}'$ means that $\mathbf{x}$ and $\mathbf{x}'$ differ by up to 1 in all four dimensions. Therefore, our mechanism is $4 \cdot \frac{1}{2} = 2$-DP.
More formally: For any $\mathbf{x} \simeq_\infty \mathbf{x}'$ there is a sequence $\mathbf{x}_0, \dots, \mathbf{x}_4$ with $\mathbf{x}_0 = \mathbf{x}$, $\mathbf{x}_4 = \mathbf{x}'$ and $\mathbf{x}_j \simeq \mathbf{x}_{j+1}$.
Because our mechanism is $\frac{1}{2}$-DP w.r.t. $\simeq$, we have:

$$\begin{aligned}
\mathbb{P}[\mathcal{M}_f(\mathbf{x}) \in Y] &= \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(0)}\right) \in Y\right] \\
&\leq e^{\frac{1}{2}} \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(1)}\right) \in Y\right] \\
&\leq e^{2 \cdot \frac{1}{2}} \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(2)}\right) \in Y\right] \\
&\dots \\
&\leq e^{4 \cdot \frac{1}{2}} \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(t)}\right) \in Y\right] = e^2 \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}'\right) \in Y\right],
\end{aligned}$$

Problem 10 (Version B) (10 credits)

0
1
2
3
4
5
6

a)

We need to solve $\max_{\mathbf{x} \simeq \mathbf{x}'} ||f(\mathbf{x}) - f(\mathbf{x}')||_1$.
Using the definition of f, this can be written as

$$\max_{\mathbf{x} \simeq \mathbf{x}'} \left|\left|\mathbf{A}\left(\mathbf{x} - \mathbf{x}'\right)\right|\right|_1$$

Due to the definition of $\simeq$, $\mathbf{x}$ and $\mathbf{x}'$ can only differ in one component and at most by 1. Thus, the above is equivalent to

$$\max_{j \in \{1,2,3,4\}} \max_{c \in [-1,1]} \left|\left|\mathbf{A} \cdot \left(c \cdot e_j\right)\right|\right|_1 ,$$

where e_j is the unit vector with non-zero entry in dimension j.
The product with the unit vector is equivalent to summing over the j'th column of A, i.e.

$$\max_{j \in \{1,2,3,4\}} \max_{c \in [-1,1]} \left|\left| c \cdot \sum_{d=1}^{4} A_{:,d} \right|\right|_1 .$$

The function evidently has its maximum at $c = \pm 1$ and thus we have

$$\Delta_1 = \max_{j \in \{1,2,3,4\}} \left|\left| \sum_{d=1}^{4} A_{:,d} \right|\right|_1 .$$

To summarize: The maximum is achieved by changing dimension corresponding to the column whose sum has the largest absolute value. In our case, the Δ_1-sensitivity is 20.

b)

0
1

$\frac{20}{\frac{1}{2}}$

c)

0
1
2
3

This holds due to group-privacy

$\mathbf{x} \simeq \mathbf{x}'$ means that $\mathbf{x}$ and $\mathbf{x}'$ differ up to 1 in exactly one dimension. $\mathbf{x} \simeq \infty \mathbf{x}'$ means that $\mathbf{x}$ and $\mathbf{x}'$ differ by up to 1 in all four dimensions. Therefore, our mechanism is $4 \cdot \frac{1}{2} = 2$-DP.

More formally: For any $\mathbf{x} \simeq_\infty \mathbf{x}'$ there is a sequence $\mathbf{x}_0, \dots, \mathbf{x}_4$ with $\mathbf{x}_0 = \mathbf{x}$, $\mathbf{x}_4 = \mathbf{x}'$ and $\mathbf{x}_j \simeq \mathbf{x}_{j+1}$.

Because our mechanism is $\frac{1}{2}$-DP w.r.t. $\simeq$, we have:

$$
\begin{aligned}
\mathbb{P}[\mathcal{M}_f(\mathbf{x}) \in Y] &= \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(0)}\right) \in Y\right] \\
&\leq e^{\frac{1}{2}} \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(1)}\right) \in Y\right] \\
&\leq e^{2 \cdot \frac{1}{2}} \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(2)}\right) \in Y\right] \\
&\dots \\
&\leq e^{4 \cdot \frac{1}{2}} \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(t)}\right) \in Y\right] = e^{2} \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}'\right) \in Y\right],
\end{aligned}
$$

Problem 10 (Version C) (10 credits)

0
1
2
3
4
5
6

a)

We need to solve $\max_{\mathbf{x}\simeq\mathbf{x}'} ||f(\mathbf{x}) - f(\mathbf{x}')||_1$.
Using the definition of f, this can be written as

$$\max_{\mathbf{x}\simeq\mathbf{x}'} ||\mathbf{A}\left(\mathbf{x} - \mathbf{x}'\right)||_1$$

Due to the definition of $\simeq$, $\mathbf{x}$ and $\mathbf{x}'$ can only differ in one component and at most by 1. Thus, the above is equivalent to

$$\max_{j\in\{1,2,3,4\}} \max_{c\in[-1,1]} ||\mathbf{A}\cdot\left(c\cdot e_j\right)||_1,$$

where e_j is the unit vector with non-zero entry in dimension j.
The product with the unit vector is equivalent to summing over the j'th column of A, i.e.

$$\max_{j\in\{1,2,3,4\}} \max_{c\in[-1,1]} \left|\left|c\cdot\sum_{d=1}^{4} A_{:,d}\right|\right|_1.$$

The function evidently has its maximum at $c = \pm 1$ and thus we have

$$\Delta_1 = \max_{j\in\{1,2,3,4\}} \left|\left|\sum_{d=1}^{4} A_{:,d}\right|\right|_1.$$

To summarize: The maximum is achieved by changing dimension corresponding to the column whose sum has the largest absolute value. In our case, the Δ_1-sensitivity is 18.

b)

0
1

$\frac{18}{\frac{1}{2}}$

c)

0
1
2
3

This holds due to group-privacy
$\mathbf{x} \simeq \mathbf{x}'$ means that $\mathbf{x}$ and $\mathbf{x}'$ differ up to 1 in exactly one dimension. $\mathbf{x} \simeq \infty \mathbf{x}'$ means that $\mathbf{x}$ and $\mathbf{x}'$ differ by up to 1 in all four dimensions. Therefore, our mechanism is $4 \cdot \frac{1}{2} = 2$-DP.
More formally: For any $\mathbf{x} \simeq_\infty \mathbf{x}'$ there is a sequence $\mathbf{x}_0, \ldots, \mathbf{x}_4$ with $\mathbf{x}_0 = \mathbf{x}$, $\mathbf{x}_4 = \mathbf{x}'$ and $\mathbf{x}_j \simeq \mathbf{x}_{j+1}$.
Because our mechanism is $\frac{1}{2}$-DP w.r.t. $\simeq$, we have:

$$\begin{aligned}
\mathbb{P}[\mathcal{M}_f(\mathbf{x}) \in Y] &= \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(0)}\right) \in Y\right] \\
&\leq e^{\frac{1}{2}} \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(1)}\right) \in Y\right] \\
&\leq e^{2 \cdot \frac{1}{2}} \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(2)}\right) \in Y\right] \\
&\ldots \\
&\leq e^{4 \cdot \frac{1}{2}} \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(t)}\right) \in Y\right] = e^{2} \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}'\right) \in Y\right],
\end{aligned}$$

Problem 10 (Version D) (10 credits)

0
1
2
3
4
5
6

a)

We need to solve $\max_{\mathbf{x}\simeq\mathbf{x}'} \|f(\mathbf{x}) - f(\mathbf{x}')\|_1$.
Using the definition of f, this can be written as

$$\max_{\mathbf{x}\simeq\mathbf{x}'} \left\|\mathbf{A}\left(\mathbf{x}-\mathbf{x}'\right)\right\|_1$$

Due to the definition of $\simeq$, $\mathbf{x}$ and $\mathbf{x}'$ can only differ in one component and at most by 1. Thus, the above is equivalent to

$$\max_{j\in\{1,2,3,4\}} \max_{c\in[-1,1]} \left\|\mathbf{A}\cdot\left(c\cdot e_j\right)\right\|_1,$$

where e_j is the unit vector with non-zero entry in dimension j.
The product with the unit vector is equivalent to summing over the j'th column of A, i.e.

$$\max_{j\in\{1,2,3,4\}} \max_{c\in[-1,1]} \left\|c\cdot\sum_{d=1}^{4} A_{:,d}\right\|_1.$$

The function evidently has its maximum at $c = \pm 1$ and thus we have

$$\Delta_1 = \max_{j\in\{1,2,3,4\}} \left\|\sum_{d=1}^{4} A_{:,d}\right\|_1.$$

To summarize: The maximum is achieved by changing dimension corresponding to the column whose sum has the largest absolute value. In our case, the Δ_1-sensitivity is 32.

b)

0
1

$\frac{32}{\frac{1}{2}}$

c)

0
1
2
3

This holds due to group-privacy
$\mathbf{x} \simeq \mathbf{x}'$ means that $\mathbf{x}$ and $\mathbf{x}'$ differ up to 1 in exactly one dimension. $\mathbf{x} \simeq \infty\mathbf{x}'$ means that $\mathbf{x}$ and $\mathbf{x}'$ differ by up to 1 in all four dimensions. Therefore, our mechanism is $4 \cdot \frac{1}{2} = 2$-DP.
More formally: For any $\mathbf{x} \simeq_\infty \mathbf{x}'$ there is a sequence $\mathbf{x}_0, \dots, \mathbf{x}_4$ with $\mathbf{x}_0 = \mathbf{x}$, $\mathbf{x}_4 = \mathbf{x}'$ and $\mathbf{x}_j \simeq \mathbf{x}_{j+1}$.
Because our mechanism is $\frac{1}{2}$-DP w.r.t. $\simeq$, we have:

$$\begin{aligned}
\mathbb{P}[\mathcal{M}_f(\mathbf{x}) \in Y] &= \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(0)}\right) \in Y\right] \\
&\leq e^{\frac{1}{2}} \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(1)}\right) \in Y\right] \\
&\leq e^{2 \cdot \frac{1}{2}} \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(2)}\right) \in Y\right] \\
&\dots \\
&\leq e^{4 \cdot \frac{1}{2}} \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}^{(t)}\right) \in Y\right] = e^{2} \cdot \mathbb{P}\left[\mathcal{M}_f\left(\mathbf{x}'\right) \in Y\right],
\end{aligned}$$

Problem 11 (Version A) (11 credits)

0
1

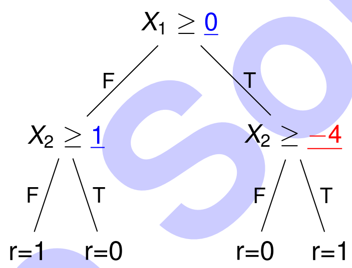
a)

ID	1	2	3	4	5	6
A	a	a	a	b	b	b
Y	0	0	1	0	1	1
R	0	1	0	0	1	1

0
1
2
3

b)

Change -2 to -4.



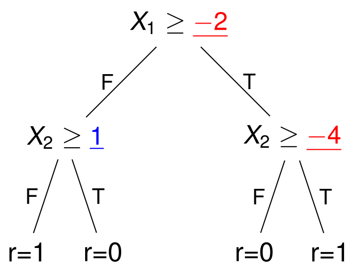
Resulting predictions:

ID	1	2	3	4	5	6
A	a	a	a	b	b	b
Y	0	0	1	0	1	1
R	0	1	1	0	1	1

c)

0
1
2
3

Change -2 to -4 and 0 to -2.



Resulting predictions:

ID	1	2	3	4	5	6
A	a	a	a	b	b	b
Y	0	0	1	0	1	1
R	0	0	1	0	1	1

0 1 2

d)

Yes, it exists. See subproblem b.).

0 1 2

e)

No, because Y and A are correlated / not independent (see homework week 14, problem 3).

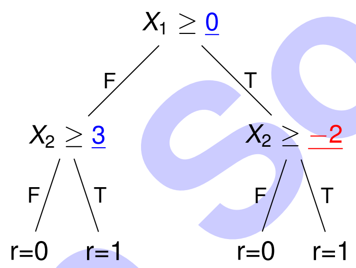
Problem 11 (Version B) (11 credits)

a)

ID	1	2	3	4	5	6
A	a	a	a	b	b	b
Y	0	1	1	0	0	1
R	0	1	1	0	1	0

b)

Change 1 to -2.



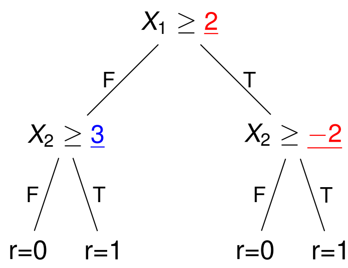
Resulting predictions:

ID	1	2	3	4	5	6
A	a	a	a	b	b	b
Y	0	1	1	0	0	1
R	0	1	1	0	1	1

0
1
2
3

c)

Change 1 to -2 and 0 to 2.



Resulting predictions:

ID	1	2	3	4	5	6
A	a	a	a	b	b	b
Y	0	1	1	0	0	1
R	0	1	1	0	0	1

d)

0
1
2

Yes, it exists. See subproblem b.).

e)

0
1
2

No, because Y and A are correlated / not independent (see homework week 14, problem 3).

Problem 11 (Version C) (11 credits)

0 1

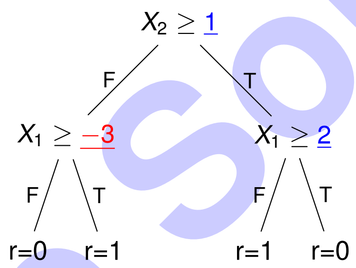
a)

ID	1	2	3	4	5	6
A	a	a	a	b	b	b
Y	0	0	1	0	1	1
R	0	1	0	0	1	1

0 1 2 3

b)

Change -1 to -3.



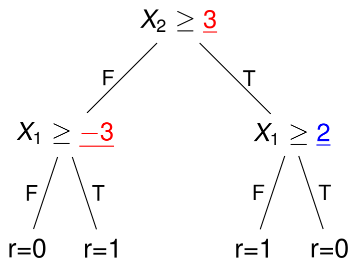
Resulting predictions:

ID	1	2	3	4	5	6
A	a	a	a	b	b	b
Y	0	0	1	0	1	1
R	0	1	1	0	1	1

c)

0
1
2
3

Change -1 to -3 and 1 to 3.



Resulting predictions:

ID	1	2	3	4	5	6
A	a	a	a	b	b	b
Y	0	0	1	0	1	1
R	0	0	1	0	1	1

0
1
2

d)

Yes, it exists. See subproblem b.).

0
1
2

e)

No, because Y and A are correlated / not independent (see homework week 14, problem 3).

Problem 11 (Version D) (11 credits)

a)

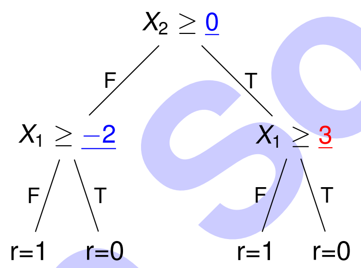
0
1

ID	1	2	3	4	5	6
A	a	a	a	b	b	b
Y	0	1	1	0	0	1
R	0	1	1	0	1	0

b)

0
1
2
3

Change child 0 to 3



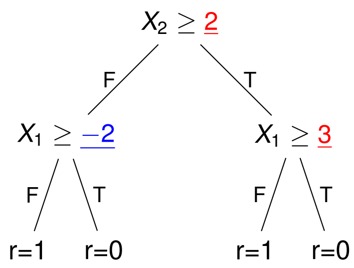
Resulting predictions:

ID	1	2	3	4	5	6
A	a	a	a	b	b	b
Y	0	1	1	0	0	1
R	0	1	1	0	1	1

0 1 2 3

c)

Change child 0 to 3 and root 0 to 2.



Resulting predictions:

ID	1	2	3	4	5	6
A	a	a	a	b	b	b
Y	0	1	1	0	0	1
R	0	1	1	0	0	1

d)

0
1
2

Yes, it exists. See subproblem b.).

e)

0
1
2

No, because Y and A are correlated / not independent (see homework week 14, problem 3).

Additional space for solutions–clearly mark the (sub)problem your answers are related to and strike out invalid solutions.

Sample Solution

Sample Solution

Sample Solution