Jean Monnet University

"Data Analysis" Exam
MLDM Master program

December 2023 - **(2 hours)**

Pocket calculators and two 2-sided papers of handwritten notes are allowed.

Exercise 1 (4.5 points): Gaussian distribution

Invented by the German psychologist William Stern, the intelligence quotient (IQ) corresponds to the rank of a person relative to a population $\mathcal{P}$ represented by a Gaussian distribution of mean $\mu = 100$ and standard deviation $\sigma = 15$.

1. What is the probability to have an $IQ > 130$?
2. What IQ corresponds to the top 1% smartest people?
3. If we randomly draw 10 people from the population $\mathcal{P}$. Let X be the number of persons among these 10 with a $IQ > 110$. Calculate the probability $P(X = 1)$.

Exercise 2 (4.5 points): Likelihood Maximization

Let $x_1, x_2, ..., x_n$ be the observations of n Gaussian variables $X_1, X_2, ..., X_n$ i.i.d. according to the following normal distribution $N(\mu, \sigma)$:

$$f_{\mu,\sigma}(x_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

1. Using the likelihood maximization, find two empirical estimates $\hat{\mu}$ and $\hat{\sigma}$ of the two parameters μ and σ respectively. For the sake of calculation, you might make use of the log-likelihood.

Exercise 3 (6 points): PCA

Let $A = (-1, -1), B = (1, 1), C = (3, -3)$ be a set 3 points in $\mathbb{R}^2$.

1. Project the data onto the line $\mathbb{R}$ by performing all the steps of a PCA.
2. Compute the part of the variance explained by the projection. What do you conclude?
3. Plot the data before and after the projection on the same figure.

Exercise 4 (5 points): Multiple Choice Questions

Circle the letter corresponding to the correct answer (**only one is correct**).
- Each correct answer **adds 1**.
- Each incorrect answer **subtracts 1/2**.

1. The determinant of the matrix $A = \begin{pmatrix} 1 & 3 & 3 \\ 2 & 0 & 2 \\ 1 & 2 & 1 \end{pmatrix}$ is:

   a. $det(A) = 4$
   b. $det(A) = 8$
   c. $det(A) = 12$

2. The inverse of the matrix $A = \begin{pmatrix} 2 & 1 \\ -1 & 3 \end{pmatrix}$ is:

   a. $A^{-1} = \begin{pmatrix} 3/7 & -1/7 \\ 1/7 & 2/7 \end{pmatrix}$

   b. $A^{-1} = \begin{pmatrix} 3/5 & -1/5 \\ 1/5 & 2/5 \end{pmatrix}$

   c. $A^{-1} = \begin{pmatrix} 3/7 & 1/7 \\ -1/7 & 2/7 \end{pmatrix}$

3. What is the covariance $COV(XY)$ between X and Y whose joint distribution is defined as follows?

$X\backslash Y$	0	1
1	1/2	0
2	0	1/2

   a. $COV(XY) = 0$

   b. $COV(XY) = 1/2$

   c. $COV(XY) = 1/4$

4. Given $f(x) = \alpha x$, $\forall x \in [0, 1]$. What is the value of α so that $f(x)$ is an actual density function?

   a. $\alpha = 1$

   b. $\alpha = 2$

   c. $\alpha = 3$

5. Let x and y be respectively the width and height of a rectangle R. Which values of x and y allow to maximize the area of R and such that the perimeter is equal to 4?

   a. $x = 1$ and $y = 1$

   b. $x = 1/2$ and $y = 3/2$

   c. $x = 3/4$ and $y = 5/4$