Data Science Preliminary Exam
(June 16th, 2025)

First name : ____________	Student ID : ____________
Last name : ____________	Additional pages : ____________

Instructions:

1. **All problems are worth 10 points.**
2. **Explain your answers clearly. Unclear answers will not receive credit.**
   State results and theorems you are using.
3. **Use first the front and back of each page to write the solution of each problem.**
4. **If you need extra pages, please <u>do not</u> use the same sheet for different problems.**
5. **Write your name and problem number on any additional page you use.**

PROBLEM 1.

(a) If $A = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}$, find the value of A^{300}.

(b) For what range of numbers a,b are the matrices A,B positive definite?

$$A = \begin{bmatrix} a & 2 & 2 \\ 2 & a & 2 \\ 2 & 2 & a \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 & 4 \\ 2 & b & 8 \\ 4 & 8 & 7 \end{bmatrix}$$

PROBLEM 2.

1. Let $A = \begin{bmatrix} -4 & -6 \\ 3 & -8 \end{bmatrix}$

   (a) Determine a singular value decomposition of A. (b) Compute the pseudo-inverse of A. (d) What is the image of the unit circle under A? (e) What is the Frobenius norm of A? (f) Write the matrix A as a sum of rank 1 matrices.

2. Let A be an $m \times n$ matrix, imagine A is written on a piece of paper. B is obtained from A by rotating 90 degrees clockwise the piece of paper. Visually, that means you are rotating your point of view of the matrix, not performing a matrix operation. For example we get

$$A = \begin{bmatrix} 1 & 2 & 3 & 0 & 5 \\ 0 & -1 & 0 & 4 & 6 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & 6 \\ 0 & 4 \\ 3 & 0 \\ 2 & -1 \\ 1 & 0 \end{bmatrix}$$

   Do A and B have the same singular values? Why?

PROBLEM 3.

1. Dualize the following linear optimization problems (use y_i for the dual variables)

$$\begin{aligned} \max \quad & 4x_1 + 2x_2 - 3x_3 \\ & 1x_1 - 2x_2 + 4x_3 \leq 5 \\ & 3x_1 + 1x_2 \leq 2 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

$$\begin{aligned} \min \quad & 4x_1 + 2x_2 - 3x_3 \\ & 17x_1 - 5x_2 + 4x_3 \geq 5 \\ & 3x_1 + 4x_2 = 2 \\ & x_1, x_2 \geq 0, \quad x_3 \in \mathbb{R} \end{aligned}$$

2. Introduce slack variables where necessary and write down the conditions of the complementary slackness theorem.

   Then, without running any algorithm, prove that the following pair of solutions for the first problem is optimal: $(x_1, x_2, x_3) = (0, 2, 0)$, $(y_1, y_2) = (0, 2)$.

PROBLEM 4. Suppose that there are n companies and n job applicants, that every company has received exactly k of those job applications, and that every applicant applied to exactly k companies. Is it possible to match every company with an applicant and every applicant to a company? Why?

PROBLEM 5.

We revisit the least-squares regression problem: Given a dataset of n data points $(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$, $1 \leq i \leq n$, the goal is to find $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$ so as to minimize:

$$RSS(a, b) = \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i^\intercal a - b)^2$$

In other words, we are trying to approximate $y_i \simeq x_i^\intercal a - b$ and the approximation error is measured by the function RSS above.

a. Rewrite the least-squares regression problem in matrix form, that is find $X \in \mathbb{R}^{n\times(d+1)}$ and $Y \in \mathbb{R}^n$ such that the problem above takes the form:

$$\min_{d\in\mathbb{R}^{d+1}} \|Xd - Y\|^2$$

and express X and Y in terms of the data points (x_i, y_i).

b. Define $f : \mathbb{R}^{d+1} \to \mathbb{R}$ by $f(d) = \|Xd - Y\|^2$ for all $d \in \mathbb{R}^{d+1}$. Compute the gradient and Hessian of f and show that f is convex.

c. Give a sufficient and necessary condition for f to be strongly convex.

In the following questions assume that the condition of part c. is satisfied.

d. Solve the equation $\nabla f(x) = 0$ and explain how you would use this to find an optimal solution to the least-squares regression problem.

e. An alternative approach to part d. above is to use gradient descent. For this, we need to solve for any direction $\delta \in \mathbb{R}^{d+1}$ the following line search problem:

$$\min_{\lambda\in\mathbb{R}} f(d + \lambda\delta)$$

Give a closed-form formula for the optimal solution to the line-search problem. The solution should be expressed in terms of X, Y, d and δ.

PROBLEM 6. Consider the optimization problem

$$\min\ 2x_1 \arctan(x_1) - \ln(x_1^2+1) + x_2^4 + (x_3-1)^2$$

subject to the conditions:

$$x_1^2 + x_2^2 + x_3^2 - 4 \leq 0, \quad x_1 \geq 0.$$

Use the Karush -Kuhn-Tucker theorem to find an optimal solution.