

Mathematics of Data Science Prelim Examination

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

Instructions:

1. Explain your answers clearly. Unclear answers will not receive credit.
2. Use the front and back of each page to write the solution of each problem.
3. If you need extra pages, please do not use the same sheet for different problems.
4. Write your name and the problem number in each additional page you use.

Problem 1**(10 pts)**

An experiment is designed to estimate two unknown offsets α and β . The three measurements are modeled as

$$b_1 \approx \alpha, \quad b_2 \approx \alpha + \beta, \quad b_3 \approx \beta,$$

and the observed vector is

$$b = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

Let $x = (\alpha, \beta)^T$.

- (a) Write the estimation problem in the form $\min_{x \in \mathbb{R}^2} \|Ax - b\|_2^2$. Construct a thin QR factorization $A = QR$ in which the diagonal entries of R are positive. Explain why the least-squares solution is unique.
- (b) Use the QR factorization to compute the least-squares solution x_{LS} , the fitted vector $\hat{b} = Ax_{LS}$, and the residual $r = b - \hat{b}$. Verify that r is orthogonal to $\text{range}(A)$.
- (c) Compute the pseudoinverse A^+ and the orthogonal projector P onto $\text{range}(A)$.
- (d) Find the singular values of A and the condition number $\kappa_2(A)$. Is this least-squares problem severely ill-conditioned? Justify your answer.

Problem 2**(10 pts)**

The columns of

$$X = \begin{bmatrix} 4 & 2 & 0 & -2 \\ 3 & 5 & -1 & 1 \end{bmatrix}$$

represent four observations in $\mathbb{R}^2$.

- (a) Compute the mean of the observations and the centered data matrix $\tilde{X}$.
- (b) Using the convention $S = \frac{1}{4}\tilde{X}\tilde{X}^T$, compute the covariance matrix S , its eigenvalues, and an orthonormal set of eigenvectors. Identify the first principal direction and describe the corresponding principal axis of the original data geometrically.
- (c) Find the best rank-one approximation X_1 to $\tilde{X}$ in the Frobenius norm. After adding the mean back, give the four corresponding approximations to the original observations.
- (d) What fraction of the total variance is retained by the first principal component? Compute both $\|\tilde{X} - X_1\|_F^2$ and $\|\tilde{X} - X_1\|_2$.

Problem 3**(10 pts)**

Let

$$H = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 7 \\ 4 \end{bmatrix},$$

and consider

$$f(x) = \frac{1}{2}x^T Hx - b^T x, \quad x \in \mathbb{R}^2.$$

- (a) Compute $\nabla f(x)$ and $\nabla^2 f(x)$. Find an orthonormal eigen-decomposition $H = Q\Lambda Q^T$, with the eigenvalues in decreasing order. Explain why f is strictly convex.
- (b) Find the minimizer x^* and the minimum value $f^* = f(x^*)$. In the coordinates $u = Q^T(x - x^*)$, write $f(x) - f^*$ in terms of u_1 and u_2 . Describe the center, principal directions, and semiaxis lengths of the level set
- $$\{x \in \mathbb{R}^2 : f(x) = f^* + 3\}.$$
- (c) For each $i = 1, 2$, consider the separate perturbation $b + \varepsilon q_i$, where q_i is the i th column of Q and the same nonzero scalar ε is used in both cases. Determine the resulting change in the minimizer. Comparing the Euclidean norms of the two changes, which eigendirection is more sensitive, and by what factor? Find $\kappa_2(H)$.
- (d) Consider gradient descent with a constant step size,

$$x^{(k+1)} = x^{(k)} - \eta \nabla f(x^{(k)}).$$

Determine all $\eta > 0$ for which the iterates converge to x^* from every initial point. Find the value of η that minimizes

$$\rho(\eta) = \max_{i=1,2} |1 - \eta \lambda_i(H)|,$$

and give the corresponding value of $\rho(\eta)$.

- (e) Use the Karush–Kuhn–Tucker conditions to solve

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \quad & f(x) \\ \text{s.t.} \quad & x_1 + x_2 \leq 1. \end{aligned}$$

Give the optimal point and the optimal Lagrange multiplier.

Problem 4**(10 pts)**

For $p \geq 0$, consider the linear program

$$\begin{aligned} (P_p) \quad & \max \quad 3x_1 + 3x_2 + px_3 \\ & \text{s.t.} \quad 2x_1 + x_2 + x_3 \leq 8, \\ & \quad \quad x_1 + 2x_2 + x_3 \leq 7, \\ & \quad \quad x_1, x_2, x_3 \geq 0. \end{aligned}$$

- (a) Write the dual linear program.
- (b) For $p = 1$, give primal and dual feasible solutions having the same objective value. What does this prove?
- (c) Use complementary slackness to show that the primal solution in part (b) is unique.
- (d) Determine the optimal solution, or the set of optimal solutions, and the optimal value of (P_p) for every $p \geq 0$. Justify your answer.

Problem 5**(10 pts)**

A centered response is fitted using two predictors. The design matrix and response vector are

$$X = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \end{bmatrix}, \quad y = \begin{bmatrix} 4 \\ 0 \\ -2 \\ -2 \end{bmatrix}.$$

No intercept is used. For $\lambda \geq 0$, define

$$\hat{\beta}_\lambda = \underset{\beta \in \mathbb{R}^2}{\operatorname{argmin}} \{ \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2 \}.$$

- (a) Compute $X^T X$ and $X^T y$, and derive $\hat{\beta}_\lambda$ as a function of λ . Find $\hat{\beta}_0$ and $\hat{\beta}_4$.
- (b) A validation set is given by

$$X_{\text{val}} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad y_{\text{val}} = \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix}.$$

Using validation mean squared error, choose among $\lambda = 0, 4, 12$.

- (c) Suppose $y = X\beta^* + \varepsilon$, where $\mathbb{E}[\varepsilon] = 0$ and $\operatorname{Cov}(\varepsilon) = \sigma^2 I_4$. Find $\mathbb{E}[\hat{\beta}_\lambda]$ and $\operatorname{Cov}(\hat{\beta}_\lambda)$.
- (d) Find the squared bias $\|\mathbb{E}[\hat{\beta}_\lambda] - \beta^*\|_2^2$ and the total variance $\operatorname{tr}(\operatorname{Cov}(\hat{\beta}_\lambda))$. Explain how increasing λ changes bias, variance, and effective model complexity, and describe the limits as $\lambda \rightarrow 0$ and $\lambda \rightarrow \infty$.

Problem 6**(10 pts)**

Consider the four labeled observations

x_i	y_i
(3, 0)	+1
(1, 1)	+1
(-3, 0)	-1
(-1, -1)	-1

- (a) Write the primal hard-margin support vector machine problem for these data.
- (b) Find the optimal $w \in \mathbb{R}^2$ and $c \in \mathbb{R}$. Justify optimality.
- (c) Give the separating line, identify the observations on the margin, and compute the geometric margin, defined as the distance from the separating line to a nearest observation.
- (d) Suppose the feature vector (1, 1) is also observed with label -1 . Explain why the hard-margin problem is infeasible. Write the soft-margin SVM formulation, and show that the slacks associated with the two copies of (1, 1) have sum at least 2.
- (e) Return to the original four observations. Explain why unregularized logistic regression has no finite minimizer for these data. Write a regularized logistic-regression objective that has a finite minimizer.