Exercise 3 — Statistical Learning and Data Analysis
Norms, Matrix Derivatives, and Regularized Regression

Note: Questions marked with $*$ are not mandatory for submission.

1. Norms

Let $x \in \mathbb{R}^n$, and define the p-norm as:

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \tag{1}$$

a. Show that equation (1) defines a norm for $p \geq 1$.

b. Show that equation (1) does not define a norm for $p < 1$.

c. Does the triangle inequality hold for $p = 0$? (Does it define a norm?)

d. Show that for $p \geq 1$, the unit ball

$$B_p := \{x \mid \|x\|_p \leq 1\}$$

is convex. In other words, if $x_1, x_2 \in B_p$, then for any $0 < \alpha < 1$,

$$\alpha x_1 + (1 - \alpha) x_2 \in B_p.$$

e. Show that for $0 \leq p < 1$, the unit ball B_p is not convex.

f. Show that $\|A\|_F^2 = \mathrm{Tr}(A^T A) = \mathrm{Tr}(A A^T)$.

g. * Provide a closed-form expression for the operator norm of a matrix $A \in \mathbb{R}^{m \times n}$ with respect to $p = 1$.

h. Provide a closed-form expression for the operator norm of a matrix $A \in \mathbb{R}^{m \times n}$ with respect to $p = \infty$.

i. * Provide a closed-form expression for the operator norm of a matrix $A \in \mathbb{R}^{m \times n}$ with respect to $p = 2$.

j. * Show that the operator norm can equivalently be defined as:

$$\|A\|_{op} = \sup_{\nu \neq 0, \nu \in V} \frac{\|A\nu\|}{\|\nu\|} = \sup_{\|\nu\|=1, \nu \in V} \|A\nu\|.$$

k. Show sub-multiplicativity for square matrices $A, B \in \mathbb{R}^{n \times n}$:

$$\|AB\|_{op} \leq \|A\|_{op} \cdot \|B\|_{op}.$$

2. Least Squares and Matrix Derivatives

a. Show that $\nabla_\beta(z^T\beta) = z$, where $z, \beta \in \mathbb{R}^{n \times 1}$.

b. Show that $\nabla_\beta(\beta^T H \beta) = 2H\beta$, where $H \in \mathbb{R}^{n \times n}$ is symmetric.

c. Given a sample $\{x_i\}_{i=1}^n \subset \mathbb{R}$, find the empirical loss minimizer:

$$c_1^* = \arg\min_{c \in \mathbb{R}} \sum_{i=1}^n |x_i - c|.$$

d. Given a sample $\{x_i\}_{i=1}^n \subset \mathbb{R}$, find the empirical loss minimizer:

$$c_2^* = \arg\min_{c \in \mathbb{R}} \sum_{i=1}^n (x_i - c)^2.$$

e. * Given a sample $\{x_i\}_{i=1}^n \subset \mathbb{R}^d$, find the empirical loss minimizer (no closed-form solution):

$$\bar{c}_1^* = \arg\min_{c \in \mathbb{R}^d} \sum_{i=1}^n \|x_i - c\|.$$

f. Given a sample $\{x_i\}_{i=1}^n \subset \mathbb{R}^d$, find the empirical loss minimizer:

$$\bar{c}_2^* = \arg\min_{c \in \mathbb{R}^d} \sum_{i=1}^n \|x_i - c\|^2.$$

3. Regularized Least Squares Regression

We define Ridge and Lasso regression:

$$\hat{\beta}_{\text{Ridge}} = \arg\min_\beta \|X\beta - y\|_2^2 + \lambda_{\text{Ridge}}\|\beta\|_2^2 \tag{2}$$

$$\hat{\beta}_{\text{Lasso}} = \arg\min_\beta \|X\beta - y\|_2^2 + \lambda_{\text{Lasso}}\|\beta\|_1 \tag{3}$$

a. Show that the closed-form solution to equation (2) is:

$$\hat{\beta}_{\text{Ridge}} = (X^T X + \lambda I)^{-1} X^T y,$$

where I is the identity matrix.

b. Provide sufficient conditions for the invertibility of $X^T X + \lambda I$.