STATS 202: Statistical Learning and Data Science

Instructor: Linh Tran

HOMEWORK # 2

Due date: July 14, 2025

Stanford University

Introduction

Homework problems are selected from the course textbook: *An Introduction to Statistical Learning*.

Problem 1 (5 points)

Chapter 4, Exercise 1 (p. 189).

Problem 2 (5 points)

Chapter 4, Exercise 4 (p. 189).

Problem 3 (5 points)

Chapter 4, Exercise 6 (p. 191).

Problem 4 (5 points)

Chapter 4, Exercise 8 (p. 191).

Problem 5 (5 points)

Chapter 4, Exercise 13 parts a-h (p. 193)

Problem 6 (5 points)

Chapter 5, Exercise 2 (p. 219).

Problem 7 (5 points)

Chapter 5, Exercise 5 (p. 220).

Problem 8 (5 points)

Chapter 5, Exercise 6 (p. 221).

Problem 9 (5 points)

Chapter 5, Exercise 8 (p. 222).

Problem 10 (5 points)

Chapter 5, Exercise 9 (p. 223).

Problem 11 (Bonus 5 points)

Let $\mathbf{Y}$ be an indicator response matrix (i.e. a matrix of binary values 0 or 1) and suppose we apply linear regression. Specifically, let $\hat{\mathbf{Y}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}}$. Similarly, for any input $x \in \mathbb{R}^p$, we get a transformed vector $\hat{y} = \hat{\boldsymbol{\beta}}^T x$. Show that applying LDA using $\hat{\mathbf{Y}}$ is identical to LDA in the original space.