STATS 202: Statistical Learning and Data Science

Instructor: Linh Tran

HOMEWORK # 3

Due date: July 28, 2025

Stanford University

Introduction

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

Problem 1 (7 points)

Chapter 6, Exercise 3 (p. 283).

Problem 2 (7 points)

Chapter 6, Exercise 4 (p. 284).

Problem 3 (7 points)

Chapter 6, Exercise 9 (p. 286). Don't do parts (e), (f), and (g).

Problem 4 (7 points)

Chapter 7, Exercise 1 (p. 321).

Problem 5 (7 points)

Chapter 7, Exercise 8 (p. 324). Find at least one non-linear estimate which does better than linear regression, and justify this using a t-test or by showing an improvement in the cross-validation error with respect to a linear model. You must also produce a plot of the predictor X vs. the non-linear estimate $\hat{f}(X)$.

Problem 6 (7 points)

Chapter 9, Exercise 1 (p. 398).

Problem 7 (8 points)

Chapter 9, Exercise 8 (p. 401).

Problem 8 (Bonus 5 points)

In class, we reviewed that the variance from bagging is

$$\rho\sigma^2 + \frac{1-\rho}{B}\sigma^2. \tag{1}$$

Derive this formula. Furthermore, this appears to fail if ρ is negative; diagnose the problem in this case.