STATS 202: Statistical Learning and Data Science

Instructor: Linh Tran

HOMEWORK # 4

Due date: August 6, 2025

Stanford University

Introduction

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

Problem 1 (12 points)

Chapter 8, Exercise 4 (p. 362).

Problem 2 (12 points)

Chapter 8, Exercise 8 (p. 363).

Problem 3 (13 points)

Chapter 8, Exercise 10 (p. 364).

Problem 4 (13 points)

Chapter 10, Exercise 3 (p. 459).

Problem 5 (Bonus 5 points)

Let $x_i : i = 1, ..., p$ be the input predictor values and $a_k^{(2s)} : k = 1, ..., K$ be the K-dimensional output from a 2-layer and M-hidden unit neural network with sigmoid activation $\sigma(a) = \{1 + e^{-a}\}^{-1}$ such that

$$a_j^{(1s)} = w_{j0}^{(1s)} + \sum_{i=1}^{p} w_{ji}^{(1s)} x_i : j = 1, ..., M$$

$$a_k^{(2s)} = w_{k0}^{(2s)} + \sum_{j=1}^{M} w_{kj}^{(2s)} \sigma\left(a_j^{(1s)}\right)$$

Show that there exists an equivalent network that computes exactly the same output values, but with hidden unit activation functions given by $tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$, i.e.

$$a_j^{(1t)} = w_{j0}^{(1t)} + \sum_{i=1}^{p} w_{ji}^{(1t)} x_i : j = 1, ..., M$$

$$a_k^{(2t)} = w_{k0}^{(2t)} + \sum_{j=1}^{M} w_{kj}^{(2t)} \tanh\left(a_j^{(1t)}\right)$$

Hint: first derive the relation between $\sigma(a)$ and $tanh(a)$. Then show that the parameters of the two networks differ by linear transformations.