Introduction to Machine Learning

Homework 2

# 1Directions:

* **Due: Thursday February 10, 2022 at 9pm.** Late submissions will be accepted for 24 hours after that time, with a 15% penalty. (the enforcement is strict, beginning at 9:01pm, except for extreme situations; having a poor wifi connection or minor computer problems is not sufficient for the penalty to be waived.)
* Upload the homework to Canvas as a single pdf file.
* If the graders cannot easily read your submission (writing is illegible, image is too dark, or if the contrast is too low) then you might receive a zero or only partial credit.
* Any non-administrative questions must be asked in office hours or (if a brief response is sufficient) Piazza.

# 2Problems

**Problem 1.** [13 points total (6,3,4)]

Book problem Chapter 3, problem 3 “Suppose we have a data set with five predictors,

*X*1 = *GPA*, . . . ”

Note: for interactions, use products, e.g. *X*4(*i*) = *GPA*(*i*) *× IQ*(*i*)

**Problem 2.** [12 points total (3 points each)]

Book problem Chapter 3, problem 4 “I collect a set of data . . . ”

**Problem 3.** [5 points]

Suppose we have a data set with one feature *X* to predict another feature *Y* . Let *n* denote

the number of samples. Let *X*¯ and *Y*¯ denote the average values of *X* and *Y* respectively in

Background pattern

Description automatically generated with low confidencethe data set

Let *β*0*∗* and *β*1*∗* denote the coefficients for the ordinary least squares (O.L.S.) solution,

A picture containing text, clock

Description automatically generated

Their values are

*.*

*∗*

Text

Description automatically generated

Using those formulas, calculate the O.L.S. model’s prediction for *X* = *X*¯

(i.e., the prediction*Y*^ for a new sample whose *X* feature has the value *X*¯ ).

**Problem 4.** [16 points total (3,3,10)]

Book problem Chapter 6, problem 1 “We perform best subset . . . ” Notes regarding the book’s pseudocode for Algorithms 6.1-6.3:

* + “RSS” stands for “residual sum of squares” which is the (un-normalized) MSE,

Text

Description automatically generated

In the book’s pseudo-codes, “RSS” refers to *training set* RSS.

* + “cross-validated prediction error” – you can read this as “Validation set MSE.”
  + “*R*2” and “adjusted *R*2” – you can ignore these for this homework