COM S 474/574 Spring 2022
Introduction to Machine Learning

Homework 2

1 Directions:

- **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.

2 Problems

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. $X_4(i) = GPA(i) \times IQ(i)$

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 $\bar{X}$ and $\bar{Y}$ denote the average values of X and Y respectively in the data set

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X(i) \qquad\qquad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y(i).$$

Let β_0^* and β_1^* denote the coefficients for the ordinary least squares (O.L.S.) solution,

$$\{\beta_0^*, \beta_1^*\} = \underset{\{\beta_0,\beta_1\}}{\arg\min} \ \frac{1}{n}\sum_{i=1}^{n}\Big(Y(i) - \big(\beta_0 + \beta_1 X(i)\big)\Big)^2.$$

Their values are

$$\beta_0^* = \bar{Y} - \beta_1^*\bar{X} \qquad\qquad \beta_1^* = \frac{\sum_{i=1}^n (X(i)-\bar{X})(Y(i)-\bar{Y})}{\sum_{i=1}^n (X(i)-\bar{X})^2}$$

Using those formulas, calculate the O.L.S. model's prediction for $X = \bar{X}$ (i.e., the prediction $\widehat{Y}$ for a new sample whose X feature has the value $\bar{X}$).

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,

$$\text{RSS} = \sum_{i=1}^{n}\Big(Y(i) - \widehat{Y}(i)\Big)^2$$

  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