**Homework 10**

**Instructions**

This homework contains **2** concepts and **3** programming questions. In MS word or a similar text editor, write down the problem number and your answer for each problem. Combine all answers for concept questions in a single PDF file. Export/print the Jupyter notebook as a PDF file including the code you implemented and the outputs of the program. Make sure all plots and outputs are visible in the PDF.

Combine all answers into a single PDF named andrewID\_hw10.pdf and submit it to Gradescope before the due date. Refer to the syllabus for late homework policy. Please assign each question a page by using the “Assign Questions and Pages” feature in Gradescope.

Here is a breakdown of the points for programming questions:

|  |  |
| --- | --- |
| Name | Points |
| M10-L1-P1 | 15 |
| M10-L2-P1 | 15 |
| M10-HW1 | 60 |

Problem 1 (5 points)

The two linear least squares regression models are fit on the same exact training and validation datasets. Below are the R2 plots for the two models. Which of the following can be said about the models?

A screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generated

(Multiple choice - choose one)

1. Model 1 is low bias but high variance, Model 2 is low variance but high bias
2. Model 1 is high bias but low variance, Model 2 is high variance but low bias
3. Model 1 is low bias but high variance, Model 2 is high bias
4. Model 1 is high bias but low variance, Model 2 is high variance

Problem 2 (5 points)

Which of the following statements is true of k-fold cross validation?

Multiple choice (choose one)

1. K-fold cross validation helps us determine how to partition the data to obtain optimal model performance
2. K-fold cross validation trains k individual models and combines their predictions to generate a better performing model
3. K-fold cross validation trains k models to find the optimal set of hyperparameters for a given dataset
4. K-fold cross validation partitions the data into k equal sized subsets, and trains k models, each time using one subset as the validation data and the rest as the training data