ACTU PS5841 Data Science in Finance & Insurance - Autumn 2021 (Y. Wang)
Assignment - 5
Assigned 10/26/21, Due 11/02/21

Problem 1. Linear Regression, Prediction Variance, Cross Validation
The file data.csv contains a sample from the known true data generating model with unkown parameters,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

A. Use code to fit the corresponding linear regression model and report the following:
a1. Regression coefficients
a2. Your estiamte of σ

B. Use code and the fitted model in A to calculate the theoretical prediction variarance associated with the test point $(x_1 = 3.14, x_2 = 3.14)$

C. Use a 5-fold cross validation approach to estimate the prediction variance of a generic test point (Hint: expected mean squared error of test sets.) Also report the mean squared errors for the test fold associated with $K = 1, ..., 5$.