

Homework 4

Read the instructions below before you start your analysis.

1. Create a Jupyter Notebook to prepare your answers. You should upload **three (3)** files on Canvas: (i) a **.ipynb** file, (ii) a **.PDF** file that is generated by exporting the output from the first file, and (iii) the **Meeting Minutes**.PDF file. The first two files should contain the required Python code, tables, charts, and all the required explanations and answers to the questions in the homework.
2. Include your group number in the name of the file you upload. For example, if your group number is 3, name the file DSCI5340_HW4_Group3.
3. **DO NOT** use an absolute directory path. I should be able to replicate your results without trying to find the input data in another directory.
4. Use a **seed of 123**, wherever necessary, to ensure replicability.
5. Label the charts and/or tables appropriately so that it is easy to understand the information contained in a chart or table.
6. Any assignment submitted after the **deadline** will be considered late and will not be graded.

Use the attached 'auto.csv' data file to build a support vector machines model to determine whether a given car gets high or low gas mileage.

1. Create a binary variable that takes on a value of 1 for cars with gas mileage above the median and 0 for cars with gas mileage below the median.
2. Fit a support vector classifier to the data using a linear kernel. Use a seed of 123, if needed. Experiment with different values of C to predict whether a car gets high or low gas mileage.
3. Report the cross-validation errors associated with different values of this parameter. Discuss the empirical implications of those results.
4. Now repeat Q2, this time using SVMs with radial and polynomial basis kernels, with different values of γ and degree and C . Discuss the empirical implications of those results.
5. Generate relevant plots to support your findings in Q2 and Q4.