CSC 295 LAB NINE

SPORTS CAR PRICE MODELS

THE ASSIGNMENT

This assignment provides practice in exploring data sets, merging data sets, performing simple and multiple regression analysis, and utilizing models.

PORSCHE AND JAGUAR DATA

You and your friend are debating whether Porsche and Jaguar sports cars maintain their value as their age and mileage increases. You'd also like to know whether the age or mileage of the car better determines the value of the car. You both gather some price data from the internet, you for a handful of Porsches, your friend for a handful of Jaguars.

The two datasets are called `Porsche.csv` and `Jaguar.csv`. Load these datasets into a fresh notebook and explore them with the tools we have discussed in class. Create simple and multiple regression models for each type of car and for the cars collectively, merging your datasets appropriately. There is no need to reserve a portion of the data for model validation.

You should form a predictive model equation for EACH car and for the cars collectively and provide predicted prices for a 10-year-old car with 55,000 miles.

EXPECTATIONS

Be sure to include the following elements in your notebook:

- Your name, date, and a description of the lab in an opening Markdown cell.
- Installation of any packages your script uses in the first cell. You should only use packages discussed in class which may include NumPy, pandas, matplotlib, and Seaborn, statsmodels, and sklearn.
- Loading the CSV data files and merging the resulting DataFrames.
- Exploration of the datasets and creation of several simple and multiple linear regression models.
- Comments in a Markdown cell at the bottom of your notebook to answer the following questions:

    1) Is mileage or age (or both) more strongly correlated to the value of a Porsche? For a Jaguar? For all the data combined? Which is the strongest correlation for each?
    2) What is the predictive equation for your **best single regression** Porsche model? What about your **best single regression** Jaguar model? What is the predictive equation for your **best single or multiple regression model** for the collective data?

3) What is the best graphical visualization to represents all your models? Plot 3-4 models on the same plot or as plots within the same figure. Be sure to title and label your plots appropriately with variable names and units.
4) What is your prediction for the values of a 10-year-old Porsche, Jaguar, and sports car with 55,000 miles according to your model equations? (3 different values) Show the math in your notebook! You may use "10" and "55000" as magic numbers but otherwise use code to access the intercept and slope values.

SUBMISSION

Submit a single Jupyter notebook. You should run and save your notebook such that it does not have to be run again to be graded.