Python Lab1 Assignment

Umais Siddiqui

Table of contents

1 Introduction

The purpose of this assignment is to review basic data manipulation and plotting in python.

2 Problem 1

2.1 Inspection of the data

In the code below I am loading the required libraries to load the file that is needed to answer Problem 1. I will also be loading the Auto data. Finally making that the missing values have been removed from the data.

```python
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

# Load data - replacing '?' with NaN and dropping them
auto = pd.read_csv('DataSets/Auto.csv', na_values='?')
auto = auto.dropna()
```

```
# (a) Identify types
# quantitative: mpg, cylinders, displacement, horsepower, weight, acceleration, year
# qualitative: origin, name
print(auto.dtypes)
```

```
mpg             float64
cylinders         int64
displacement    float64
horsepower      float64
weight            int64
acceleration    float64
year              int64
origin            int64
name                str
dtype: object
```

2.2 a. Which of the predictors are quantitative, and which are qualitative?

Quantitative: mpg, cylinders, displacement, horsepower, weight, acceleration, year.

Qualitative: origin, name. (Note: cylinders and origin are integers but often treated as categorical; however, for this context, origin is strictly qualitative/categorical).

2.3 b. What is the range of each quantitative predictor? You can answer this using the min() and max() methods in numpy.

```
# (b) & (c) Summary stats for quantitative predictors
quant_cols = ['mpg', 'cylinders', 'displacement', 'horsepower', 'weight', 'acceleration', 'ye
min_max = auto[quant_cols].agg(['min', 'max'])
print(min_max)
```

```
      mpg  cylinders  displacement  horsepower  weight  acceleration  year
min   9.0          3          68.0        46.0    1613           8.0    70
max  46.6          8         455.0       230.0    5140          24.8    82
```