Kholby Lawson
CS544
AP1

Results of linear regression using four features (displacement, horsepower, weight, and acceleration:

$$y=45.251-.006\,x_{displacement}-.044\,x_{horsepower}-.005\,x_{weight}-.023\,x_{acceleration}$$

$$R^2=0.707$$

Linear regression using displacement only:

$$y=35.121-.060\,x \quad , \quad R^2=0.648$$

Linear regression using horsepower only:

$$y=39.936-.158\,x \quad , \quad R^2=0.606$$

Linear regression using weight only:

$$y=46.217-.008\,x \quad , \quad R^2=0.693$$

Linear regression using acceleration only:

$$y=4.833+1.198\,x \quad , \quad R^2=0.179$$

Of the four predictors used above, weight is the most effective at predicting MPG. Its coefficient is about -0.0075, which can be interpreted as indicating that a 1 pound reduction in weight would correlate with a 0.0075 MPG increase in fuel economy. The coefficient of determination for the model containing all four predictors is about 0.71. This can be interpreted as saying that 71% of the variance in the response is explained by the variables in this model. We should still be able to use linear regression even after including the number of cylinders in our feature set. Furthermore, these values have an ordered relationship, so the values should not need to be modified to work correctly as the market area features were in the example. With this approach, the code is very similar to each of the approaches above, and the resulting model has a coefficient of determination of 0.708, which is slightly higher than any of the previous models, indicating that this models explains just a bit more of the variance.

```
features = ['Displacement','Horsepower', 'Weight', 'Acceleration', 'Cylinders']
X = data[features]
linreg.fit(X,y)
```

Linear regression using the four predictors above, plus the number of cylinders:

$$y=46.264-.00008\,x_{displacement}-.045\,x_{horsepower}-.005\,x_{weight}-.029\,x_{acceleration}-.040\,x_{cylinders}$$

$$R^2=0.708$$