CS544 AP1 Report – Chris Freeman

1. Which of the four predictors is the most effective at predicting MPG? You can answer this naively by regressing with each predictor alone and noting the R2 value of the model (variance explained by the model).

Looking at each individual R2 value, Weight is the most effective at predicting MPG at 0.692630433121. The R2 value for displacement was 0.648229400319. The R2 value for horsepower was 0.605948257889. The R2 value for acceleration was 0.179207050156.

1. In the single-predictor model of the best predictor, what is the coefficient on that predictor? How should you interpret it?

The coefficient of Weight is -0.06005143. This means that if weight(X) differed by one unit, mpg(Y) will differ by -0.06005143 units, on average. This means as weight increases mpg will drop.

1. In the linear regression model containing all four predictors, what is its coefficient of determination? How should you interpret that?

The coefficient of determination (R2) is 0.706981186572. This means that the combination of all four predictors (X1, X2, X3, X4 ) will give a more accurate answer for mpg (Y).

1. If we wanted to include the cylinders information from the dataset into our model, could we still use linear regression? If so, how would we do it? Show the code for your approach.

I believe we could include cylinders information into our model. The code would be as follows.

feature\_cols = [‘cylinders’,'displacement','horsepower','weight','acceleration']

X = data[feature\_cols]

y = data.mpg

# instantiate and fit

lm5 = LinearRegression()

lm5.fit(X, y)

# print the coefficients

print lm5.intercept\_

print lm5.coef\_

print lm5.score(X, y)

grid = sns.pairplot(data, x\_vars=[‘cylinders’,'displacement','horsepower','weight','acceleration'], y\_vars='mpg', kind='reg')

grid.savefig('combined.png')