

## CSC570E Machine Learning Homework 3

### Linear Regression and Decision Trees (due by the end of the day on Wednesday, July 17th)

In this homework, you need to implement a linear regression model and a decision tree.

1. Load Auto.csv data set. This is the dataset you used in Homework 1.
2. Horsepower has missing values. Convert horsepower to int and replace the missing values with the mean (rounded to int) of horsepower.
3. Use the `lm()` function to perform a multiple linear regression with *mpg* as the response variable and all other variables except *name* as predictors.
4. Which predictors appear to have a statistically significant relationship to the response?
5. What does the coefficient for the year variable suggest?
6. Write down the linear regression equation for the model.
7. Using the linear regression model, what would be the predicted value for mpg for a car with the following features:  
cylinders: 4  
displacement: 100  
horsepower: 50  
weight: 3000  
acceleration: 15  
year: 75  
origin: 1
8. Fit a C5.0 tree to the training data, with *origin* as the response variable and all the other variables except *name* as predictors. Use a random sample of 300 rows as training set, and the remaining 97 rows as testing set.
9. What is the decision tree size?
10. Draw a confusion matrix and calculate the error rate.
11. Using the decision tree, what is the predicted value of *origin* for the following car:  
mpg: 23  
cylinders: 4  
displacement: 100  
horsepower: 50  
weight: 3000  
acceleration: 15  
year: 75

Your submission must consist of two text files:

- a text file, `description.txt`, no longer than a page: Your answers to the questions above.
- a script with history of your session. Save the session into `script.Rhistory` using:  
`> savehistory("script.Rhistory")`

It is important that your script is clean, i.e., it does not contain any “drafts” or “debugging”. You can manually edit the script in order to remove any unnecessary commands, such as trials and errors. The order of the commands must follow the order stated above. Let me know if you have questions.