BMI 6950: Applied Machine Learning *Due: 24 Jan 2017*

HW2: Linear Regression

In this exercise we will explore Linear Regression learning methods. There are three text files, one containing a single feature (X.txt), and the other containing labels (y.txt). These two files are your training data. The third file contains (predict.txt) which you will provide predictions on once you construct your hypothesis. You are to use these data files for the following exercises.

1. Implementing Gradient Descent

a. Provide a scatter plot of your data. What can you say about your data from the plot?

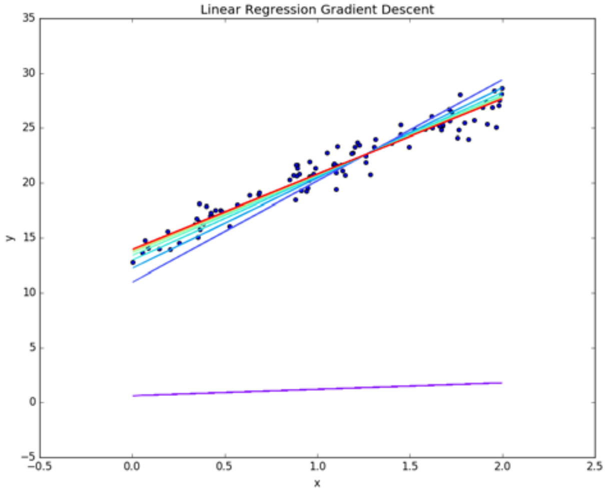
b. You will now implement the gradient decent algorithm in Python to calculate your model parameters. You can use the algorithm provided in the class lecture notes or the linear algebra version in the book. Run the algorithm for 2000 iterations. You will make multiple runs with learning rates {0.0001, 0.01, 0.5}.

   i. For each learning rate

      1. Report the model parameters ($\theta's$) for the first 5 descent iterations and the final iteration. Graph the state of the hypothesis every 200th iteration Describe what seems to be happening for each learning rate. This can be done with something like the following code statements

```
color=iter(cm.rainbow(np.linspace(0,1,10)))
...
if ((iteration % 200) == 0):
       c=next(color)
       plt.plot(X, h_theta(X), color=c)
...
plt.scatter(X,y)
```

The graph should look something like…



ii. Report your hypothesis function ($h_\theta(x) = ?$) for learning rate 0.01

iii. Predict $\hat{y}$ for the x values in the predict.txt file on your model produced with learning rate 0.01. Report your results.

iv. Provide all your working code.

2. Implementing the Scikit-Learn Linear Regression Function

a. Now use the Scikit-Learn Linear Regression function.

  i. Report your model parameters.

  ii. Report your hypothesis function ($h_\theta(x) = ?$)

  iii. Predict $\hat{y}$ for the x values in the predict.txt file. Report your results.

  iv. Provide all your working code.