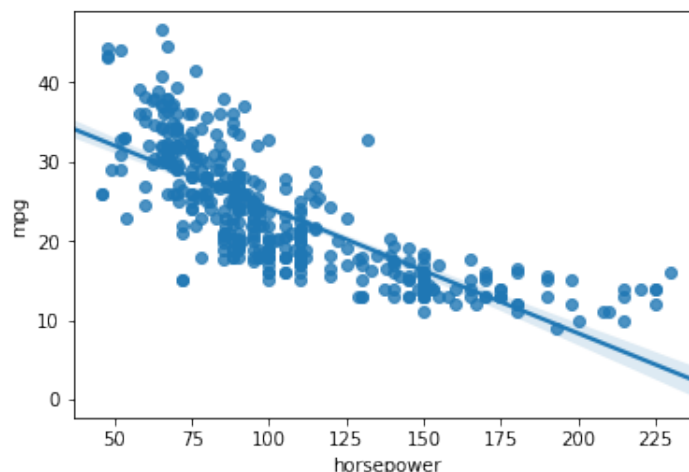**Homework 3: Due Tue 09-11-2018**
Total Points (23 pts)

1. (2 pts) (Dataset Size and the Normal Equations) Which would be more work to solve: (a) solve normal equations for a dataset having 100 features and 1000 data points (b) solve normal equations for a dataset having 1000 features and 100 data points. Justify your answer.

2. (6 pts) Read How To Launch Your Career In Deep Learning Research[1]. Give three distinct pieces of advice (don't just list three courses) from Yann LeCun for undergraduate students interested in Deep Learning research.

3. (5 pts) (Feature Engineering) Consider the height vs weight data set for children. Heights, x_1, are specified in inches and weights, y, are specified in pounds. Let us do some feature engineering. Engineer a "new" feature, x_2, equal to heights specified in centimeters and use x_2 as an additional input. Mathematically show that the minimum RMSE of the network using the engineered feature, x_2, will be equal to the minimum RMSE of the original network using only x_1. Note that 1 in = 2.54 cm.

   $$\text{original network:} \quad y = wx + b$$
   $$\text{engineered network:} \quad y = w_1 x_1 + w_2 x_2 + b$$

4. (10 pts) (Training vs Testing RMSE) Load the dataset `HP_vs_MPG.csv` into a Pandas dataframe. Split the data into 80% training and 20% testing datasets. Fit a constant, linear and parabolic function to the data and fill in the table below. (You may estimate the RMSE for the constant function by computing the standard deviation of the target values. Strictly speaking, you should use the constant determined by the training data to compute the RMSE for the test set.)

curve	training RMSE	testing RMSE
constant		
line		
parabola		

   Which function performs best at predicting MPG. Justify your answer.

---

[1]http://www.forbes.com/sites/quora/2016/08/12/how-to-launch-your-career-in-deep-learning-research/#3f26cbc32409