1. Example of regression with one variable. Consider the following simple data set of four points:
   (x, y): (1, 1), (1, 3), (4, 4), (4, 6).
   (a) Suppose you had to predict y without knowledge of x. What value would you predict? What would be its mean squared error (MSE) on these four points?
   (b) Now let's say you want to predict y based on x. What is the MSE of the linear function y = ax on these four points?
   (c) Find the line y = ax + b that minimizes the MSE on these points. What is its MSE?

2. (Code) Linear Regression: In this problem, you will analyze potential relationships between baseball player salaries and the players' statistics. You have several features: batting average, on base, runs, hits, doubles, triples, home runs, RBI, walks, strike outs, and stolen bases. Your goal will be to see if any **one** of these listed features makes for a good linear relationship. Report your linear models for each feature, the overall error in the model, and the feature (if any) that you would advice is the best **single** indicator for salary.

3. (code) In this problem, you are going to build a logistic regression digit classifier using the same MNIST dataset you used in the past homework. If you don't already have the files, first go and download them from the class website. (For a description of the data format, please see the last Homework) Please append your final code for this problem to the end of your submission. Since the label set for this task consists of ten possible digits, you will be building a multi-class logistic regression classifier. Your submission file should be in the following format: For each of the 10k images in the test_mnist.csv set, output a single line containing the ImageId and the digit you predict. You will also submit a short 1 to 2 (max) page writeup of how you evaluated the performance of your logistic regression model. Compare and contrast your kNN MNIST classifier model to your Logistic Regression MNIST classifier model.