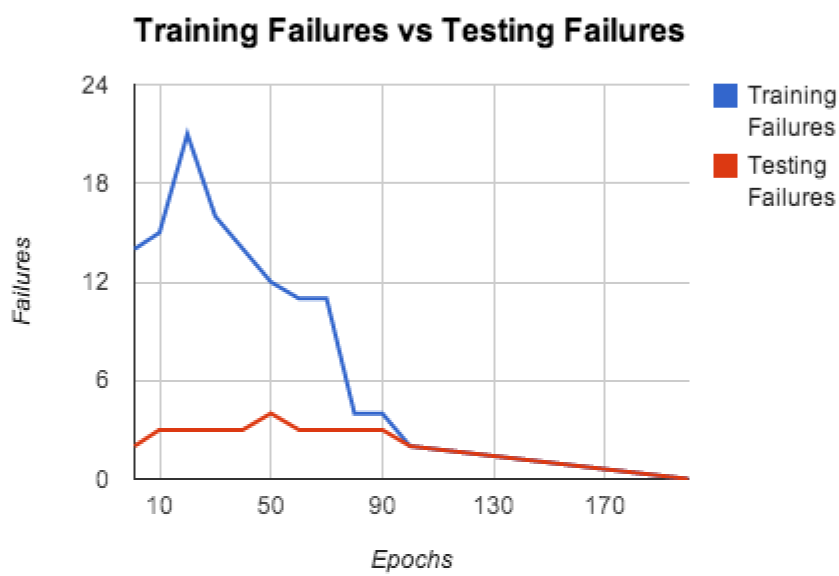
CS 434 Perceptron Algorithm

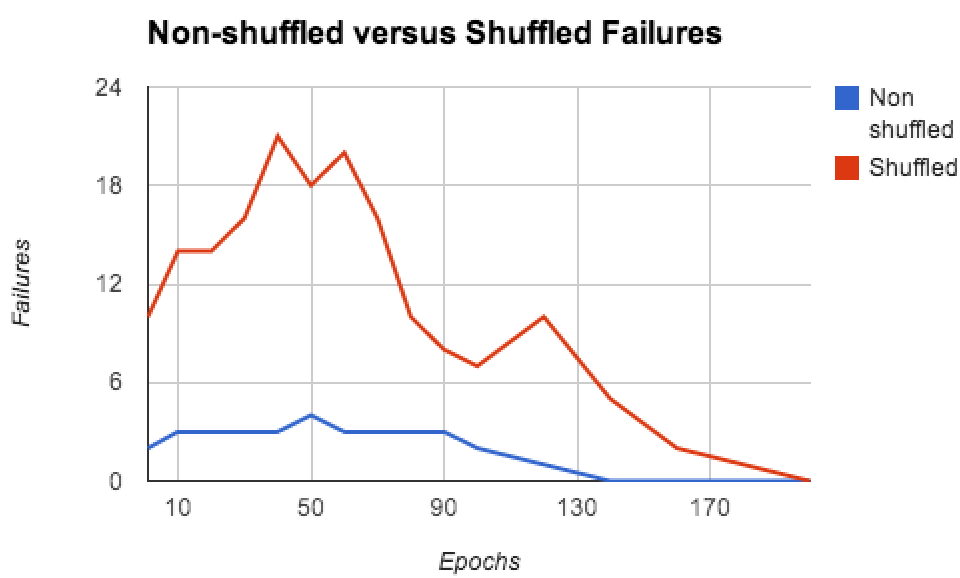
Ryley Herrington

Part I:

1. A plot of the classification accuracy of the learned classifier on both the training set and test set respectively as a function of the number of epochs (up to 100 epochs).



2. Please plot the results for shuffled and non-shuffled methods. What difference do you observe? Provide an explanation for this.



You'll notice that shuffled has many more failures.They both have similar shapes in that the more data we read in the fewer failures we observe, but still there is a large disparity between the two. This is because, as you explained in the question, that the ordered data will provide a convergence much quicker than if we just read the data in any order. In real life data that hasn't been cultivated by a teacher (assuming) then we would get many more failures and not get the correct w vector fairly quickly after we start our algorithm. But since we do shuffle it every time we could read in the outliers first and cause our line to go askew for many iterations until we finally figure out that our w should be different.

3. If a magic crystal ball tells you that the data only has two useful features, which of the three features do you think is not useful? What is your strategy for identifying it?

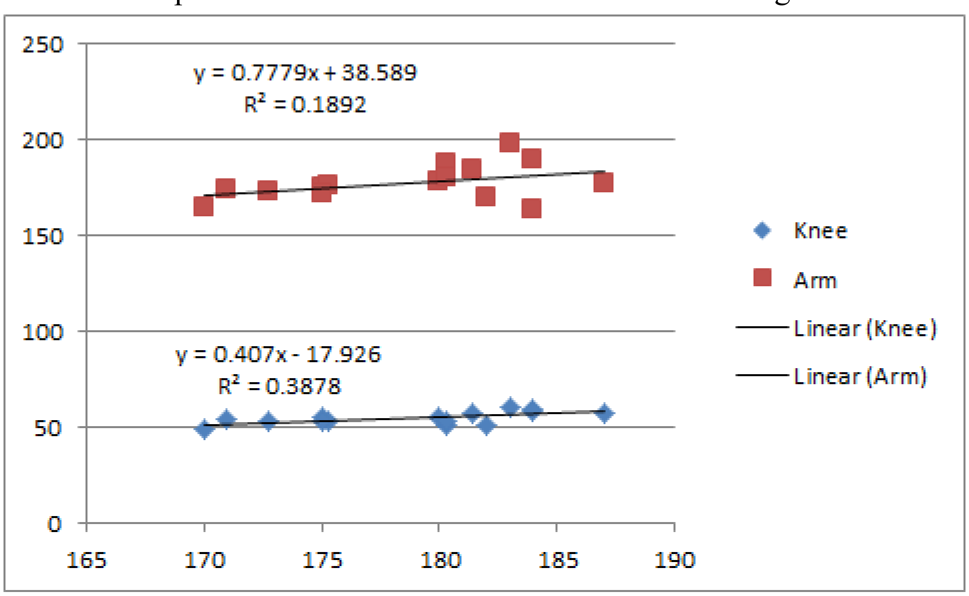
I think that if we knew only two of the data sets were useful it would be the ones with the largest deltas between data that seem to have a similar trend line to another data set. Meaning that they have some correlation; and when I graphed the data points all on the same x-axis, the first column (ie x1) barely changed with anything. The average delta between any two points was -.2, whereas the the x2 and x3 columns had growth areas in the same places and their overall curves were much more similar over 200 data points.

Part II:

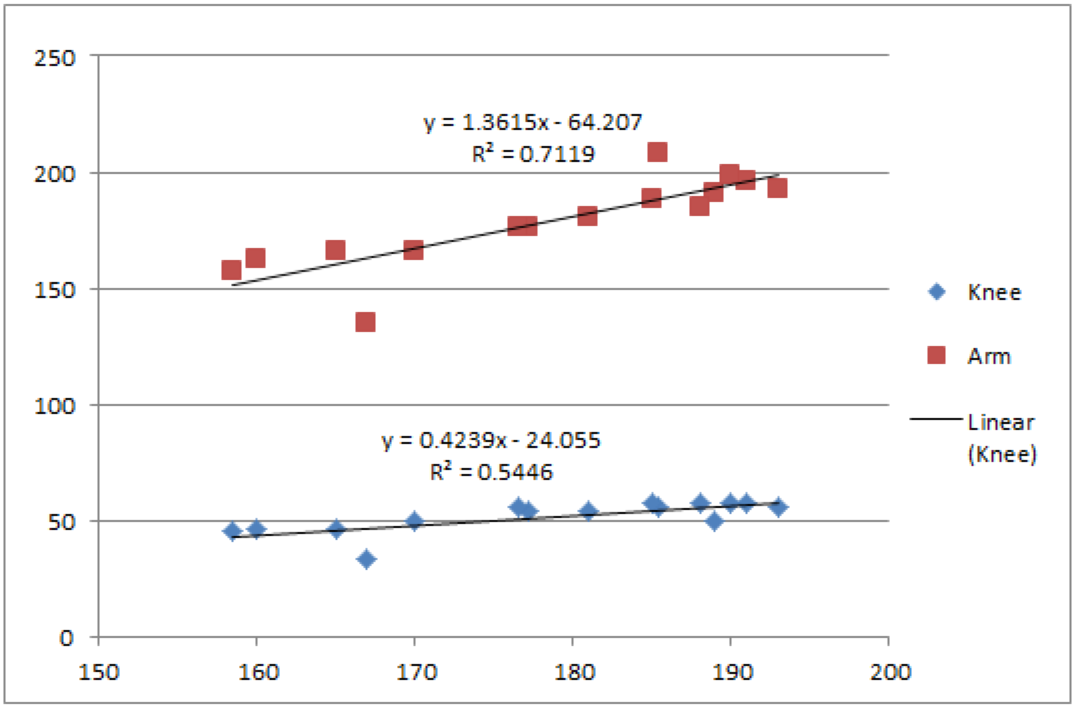
a.Please report the optimal w and b, and the resulting SSE (Sum of squared error) on the training data and test data respectively. Please also provide a plot of the training and testing data (in different colors) along with the learned regression line.

Test Data: Optimal w and b is: 0.779x + 38.589 for Armspan with an SSE of .1892

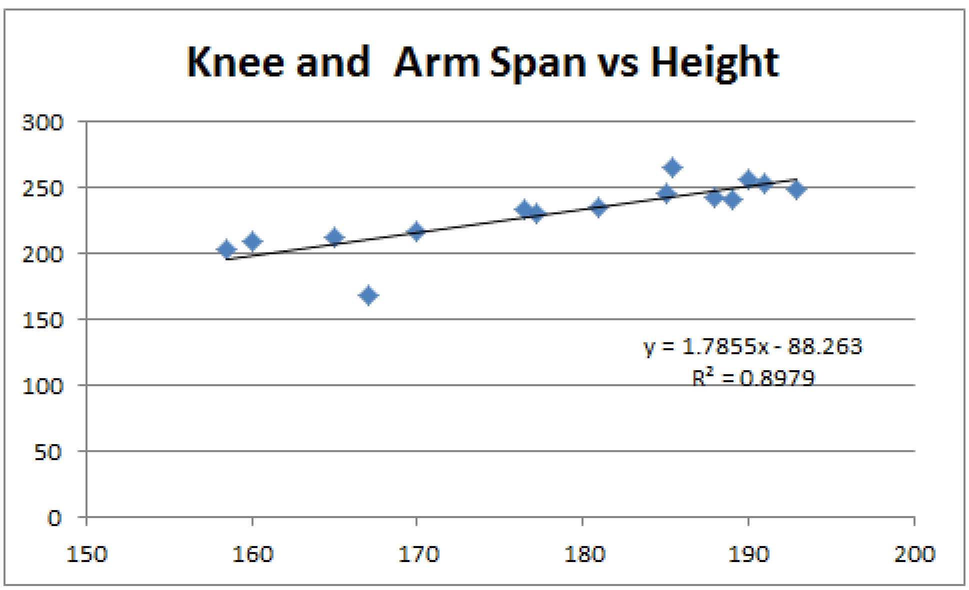
Optimal w and b is .0407x - 17.926 for knee height with an SSE of .3879



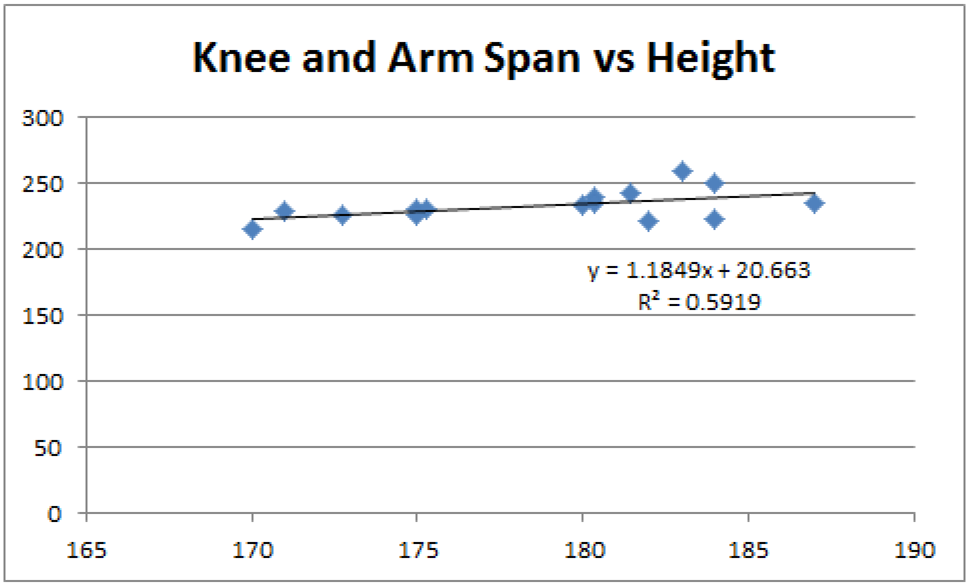
Training Data: Optimal w and b is: 1.3615x + 64.207 for Armspan with an SSE of .7119
Optimal w and b is .4239x - 24.055 for knee height with an SSE of .5446



b. Both Knee and Arm Span Training: together to determine height led to
w = 1.7855x - 88.263 and an SSE of = 0.8979



Both Knee and Arm Span Test: together to determine height led to
w = 1.1849x + 20.663 and an SSE of = 0.5919



c. When comparing all of the test data you can see that the arm span and the knee height together to learn the lsrl is much better (SSE was highest around .89) than when we do them separately. This leads me to believe that usually the training data will be better than the testing data because the SSE of training is much better than the testing data. This concurs with what we learned in the previous part as well. The training data was cultivated correctly to show that there is a correlation between arm span and knee height to height. It also makes sense because usually people are semi-proportional by just looking at them. So generally training data will have a higher SSE than testing data until we get many many test data points to learn from. Training data is probably averages of hundreds of people, whereas testing is a small subset used to represent a larger group. So of course testing data has a smaller SSE.

The one feature versus two features actually greatly helped the SSE come to a closer line. By a large margin you can tell that the new line is better to predict the height. I usually expect knee heigh to be less telling than arm span because basic bio explains that arm span is usually portional to your height(usually the same length); but knee height turned out to be a fair predictor to some degree as well. But I expect arm span to be better. Based on the testing data, knee height had a better correlation, which makes me wonder about my results. The training data that we used, and as I mentioned is probably more correct than testing, supports my claim that arm span is better to predict height. You can see that the SSE for arm span is much higher than the SSE for knee height.