AI Project 2

Jonathan Samson, Brandon Chin

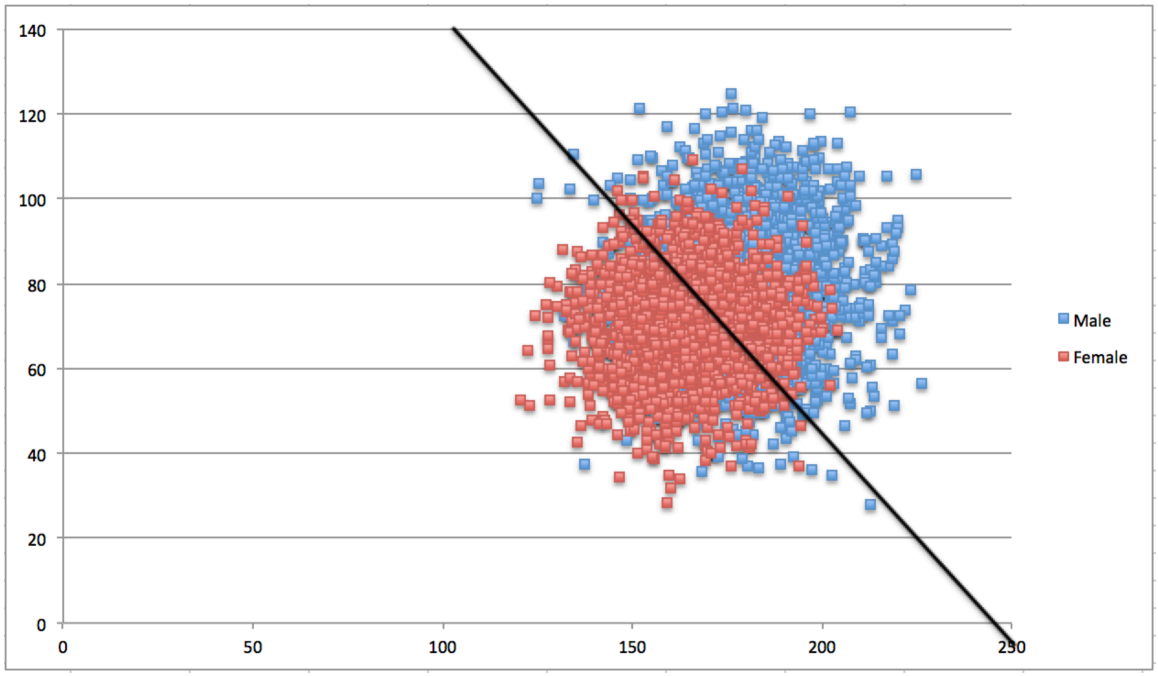
September 27, 2018

Scenario A:

Hard activation function

1. Choose 75% of the data for training, and the rest for testing for scenarios. Train and test your neuron. Plot the data and resulting separation line. Specify corresponding errors.

Plot:

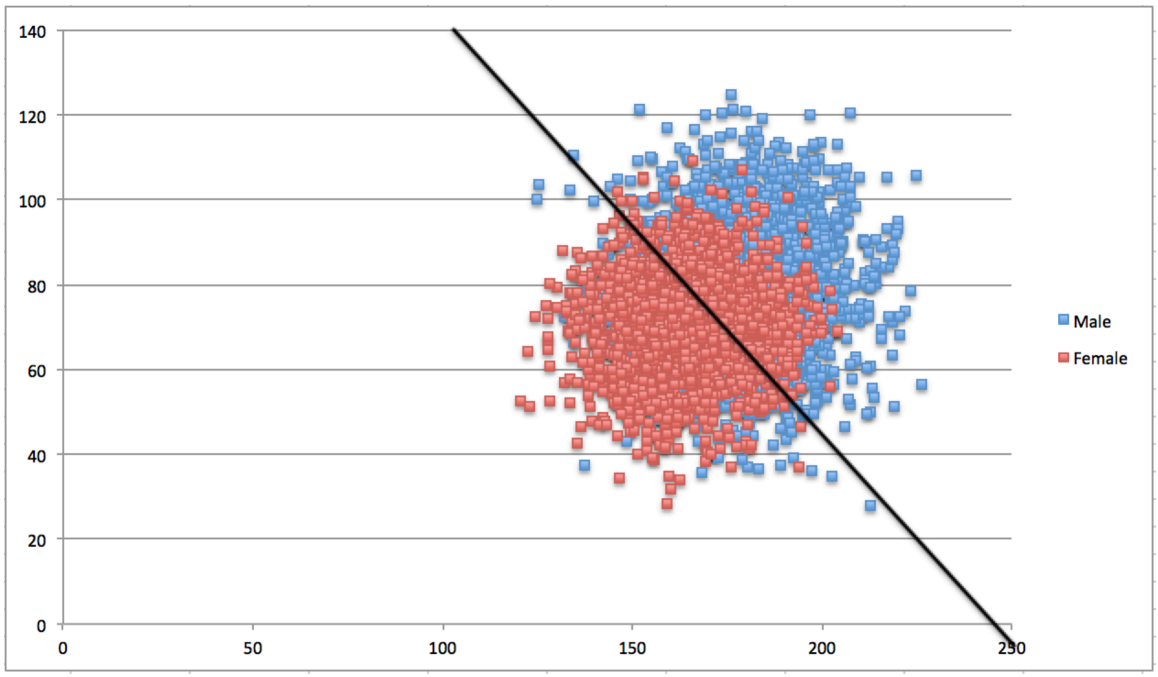


Separation Line: The equation of the separation line is y=-(0.81031x-200)/0.80477. Unfortunately Excel would not allow us to draw this line on the chart, but based off the points from a graphing calculator the separation line has been drawn.

Analysis: Out of 4000 sets of data, 3000 were randomly chosen for training, and 1000 have been chosen for testing. Of the 1000, 348 males were guessed correctly, while 135 males were wrongly guessed, and 384 females were guessed correctly, while 133 females were wrongly guessed. We had an accuracy of 73.2%. Though on every run there were slight variations in accuracy, our accuracy remained around 73%.

2. Choose 25% of the data for training, and the rest for testing for scenarios. Train and test your neuron. Plot the data and resulting separation line. Specify corresponding errors.

Plot:



Separation Line: The equation of the separation line is y=-(0.79801-200)/0.82571. Unfortunately Excel would not allow us to draw this line on the chart, but based off the points from a graphing calculator the separation line has been drawn.

Analysis: Out of 4000 sets of data, 1000 were randomly chosen for training, and 3000 have been chosen for testing. Of the 3000, 1073 males were guessed correctly, while 430 males were wrongly guessed, and 1126 females were guessed correctly, while 371 females were wrongly guessed. We had an accuracy of 73.3%. Though on every run there were slight variations in accuracy, our accuracy remained around 73%.

3. Compare 1 and 2 and discuss.

Comparing using 75% of the data compared to 25% of the data yielded very similar accuracies with the first test (75%) having an accuracy of 73.2% while the second scenario (25%) had an accuracy of 73.3%. For males only, test 1 (75%) had an accuracy of 72.05%, while test 2 (25%) had an accuracy of 71.4%. For females only, test 1 had an accuracy of 74.27%, while test 2 had an accuracy of 75.22%. The accuracies are relatively close, but from our testing we found that the accuracy did not show substantial evidence that accuracy improved with more training data compared to less training data. With at least 1000 training points (test 2) there is

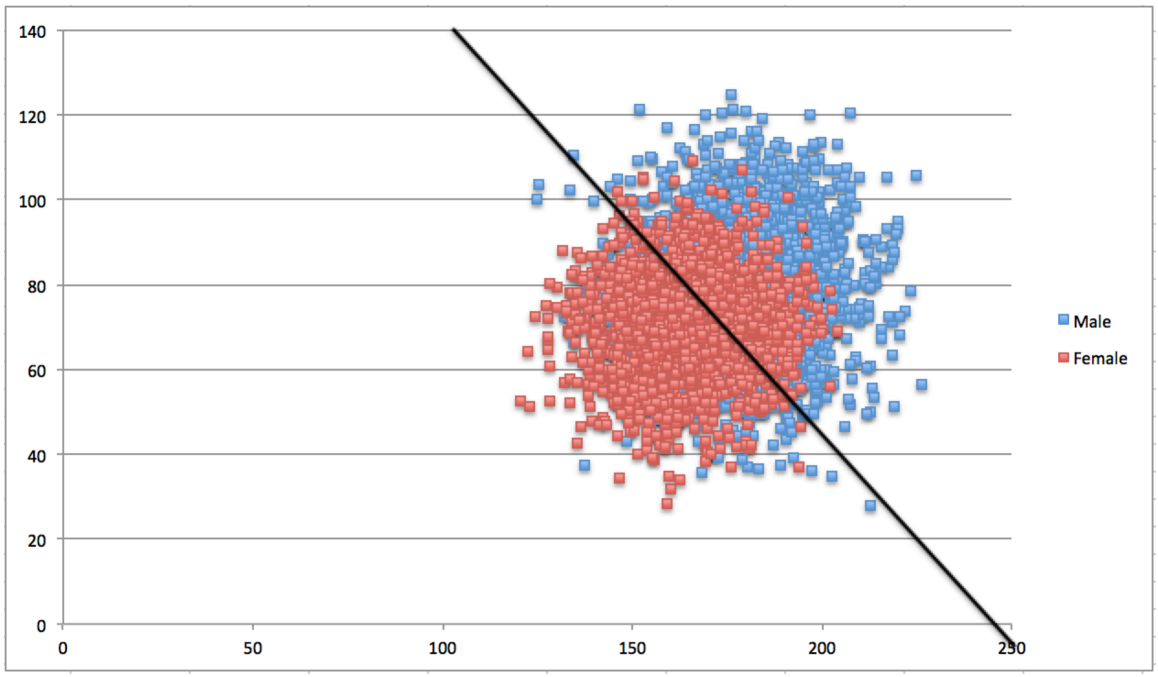
enough data to train from to achieve an accuracy that is near the accuracy of scenario 1. Additionally, having 1000 iterations helps increase the accuracy of the guesses. When the number of iterations was decreased during testing, we noticed significant decreases in the accuracy of the guesses, with the accuracy being closer to 50%.

Scenario B:

Soft activation function

1. Choose 75% of the data for training, and the rest for testing for scenarios. Train and test your neuron. Plot the data and resulting separation line. Specify corresponding errors.

Plot:



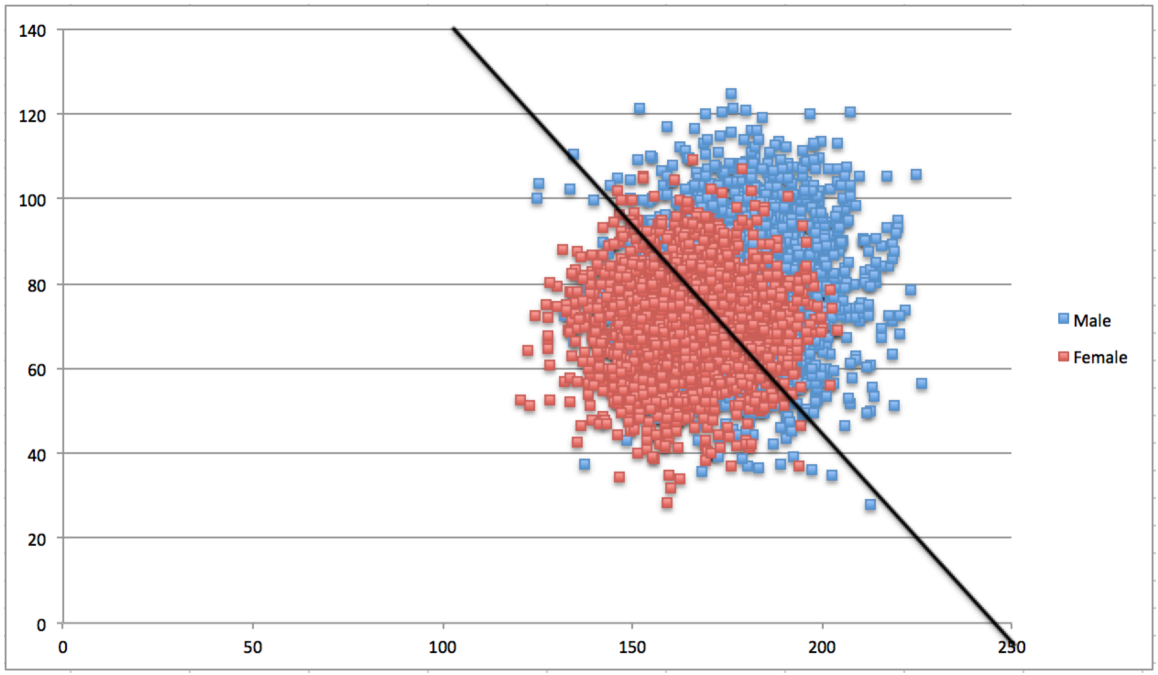
Separation Line: The equation of the separation line is y=-(0.81251-200)/0.80568. Unfortunately Excel would not allow us to draw this line on the chart, but based off the points from a graphing calculator the separation line has been drawn.

Analysis: Out of 4000 sets of data, 3000 were randomly chosen for training, and 1000 have been chosen for testing. Of the 1000, 354 males were guessed correctly, while 129 males were wrongly guessed, and 381 females were guessed correctly, while 136 females were wrongly guessed. We had an accuracy of 73.5%. Though on

every run there were slight variations in accuracy, our accuracy remained around 73%.

2. Choose 25% of the data for training, and the rest for testing for scenarios. Train and test your neuron. Plot the data and resulting separation line. Specify corresponding errors.

Plot:



Separation Line: The equation of the separation line is y=-(0.79778-200)/0.83014. Unfortunately Excel would not allow us to draw this line on the chart, but based off the points from a graphing calculator the separation line has been drawn.

Analysis: Out of 4000 sets of data, 1000 were randomly chosen for training, and 3000 have been chosen for testing. Of the 3000, 1085 males were guessed correctly, while 418 males were wrongly guessed, and 1107 females were guessed correctly, while 390 females were wrongly guessed. We had an accuracy of 73.06%. Though on every run there were slight variations in accuracy, our accuracy remained around 73%.

3. Compare 1 and 2 and discuss.

Comparing using 75% of the data compared to 25% of the data yielded very similar accuracies with the first test (75%) having an accuracy of 73.5% while the second test (25%) had an accuracy of 73.06%. For males only, test 1 had an accuracy of 73.29%, while test 2 had an accuracy of 72.19%. For females only, test 1 had an accuracy of 73.69%, while test 2 had an accuracy of 73.95%. The accuracies are

relatively close, but from our testing we found that the accuracy did not show substantial evidence that accuracy improved with more training data compared to less training data. With at least 1000 training points (test 2) there is enough data to train from to achieve an accuracy that is near the accuracy of scenario 1. Additionally, having 1000 iterations helps increase the accuracy of the guesses. When the number of iterations was decreased during testing, we noticed significant decreases in the accuracy of the guesses, with the accuracy being closer to 50%.

Scenario C:

Comparing hard and soft activation function.

1. Choose 75% of the data for training, and the rest for testing for scenarios. Train and test your neuron. Plot the data and resulting separation line. Specify corresponding errors.

For the hard activation function, out of 4000 sets of data, 3000 were randomly chosen for training, and 1000 have been chosen for testing. Of the 1000, 348 males were guessed correctly, while 135 males were wrongly guessed, and 384 females were guessed correctly, while 133 females were wrongly guessed. We had an accuracy of 73.2%.

For the soft activation function, out of 4000 sets of data, 3000 were randomly chosen for training, and 1000 have been chosen for testing. Of the 1000, 354 males were guessed correctly, while 129 males were wrongly guessed, and 381 females were guessed correctly, while 136 females were wrongly guessed. We had an accuracy of 73.5%.

We can see that there is a slight difference in accuracy with the soft activation function having a slightly better accuracy, but that is very dependent on each run. When we ran this multiple times, the percentages fluctuated and sometimes soft would be more accurate than hard and vice versa. If the number of iterations is lowered however, the hard activation function will continue to be accurate, but the soft activation function becomes less accurate as the number of iterations decreases.

2. Choose 25% of the data for training, and the rest for testing for scenarios. Train and test your neuron. Plot the data and resulting separation line. Specify corresponding errors.

For the hard activation function, out of 4000 sets of data, 1000 were randomly chosen for training, and 3000 have been chosen for testing. Of the 3000, 1073 males were guessed correctly, while 430 males were wrongly guessed, and 1126 females were guessed correctly, while 371 females were wrongly guessed. We had an accuracy of 73.3%.

For the soft activation function, out of 4000 sets of data, 1000 were randomly chosen for training, and 3000 have been chosen for testing. Of the 3000, 1085 males were guessed correctly, while 418 males were wrongly guessed, and 1107 females were guessed correctly, while 390 females were wrongly guessed. We had an accuracy of 73.06%.

Again, we can see that there is a slight difference in accuracy with the hard activation function having a slightly better accuracy, but that is very dependent on each run. When we ran this multiple times, the percentages fluctuated and sometimes soft would be more accurate than hard and vice versa. If the number of iterations is lowered however, the hard activation function will continue to be accurate, but the soft activation function becomes less accurate as the number of iterations decreases. Also, with fewer numbers of iterations, we noticed soft activation function trained more slowly, and it would sometimes not finish if number of iterations were too low.