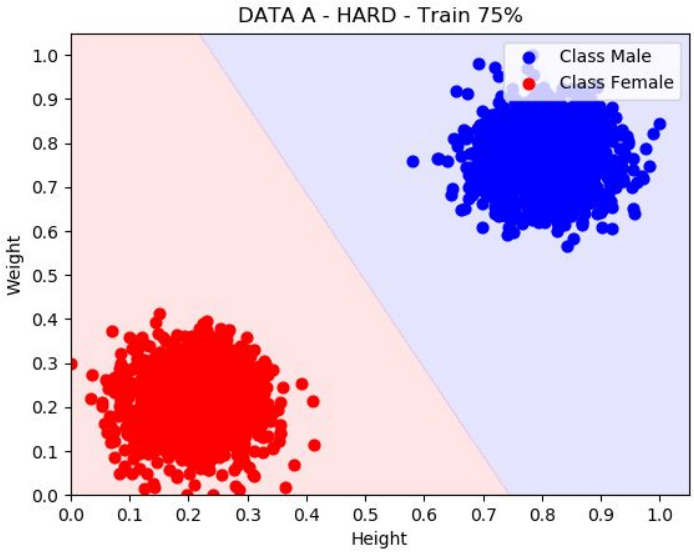
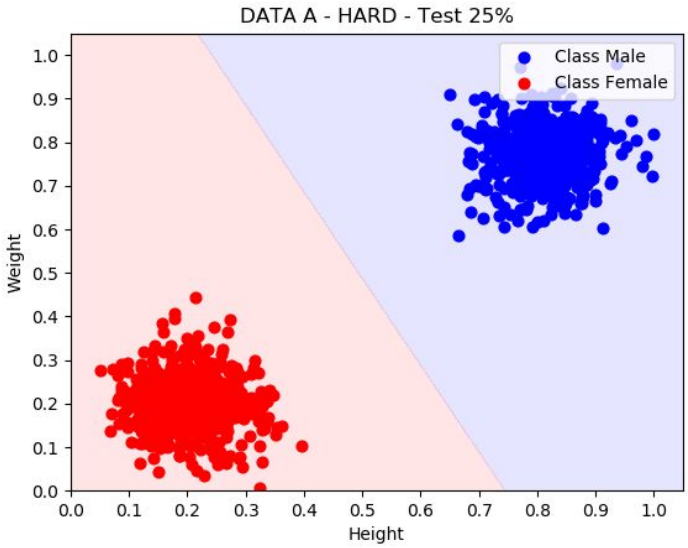
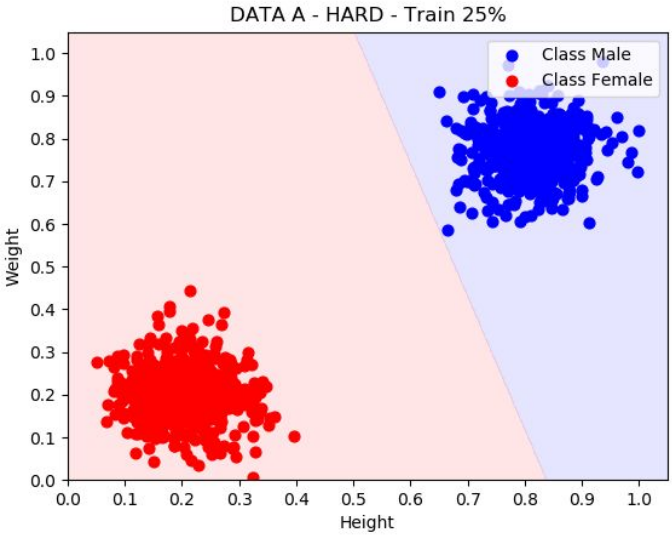
2.1 Group A Hard Activation Function

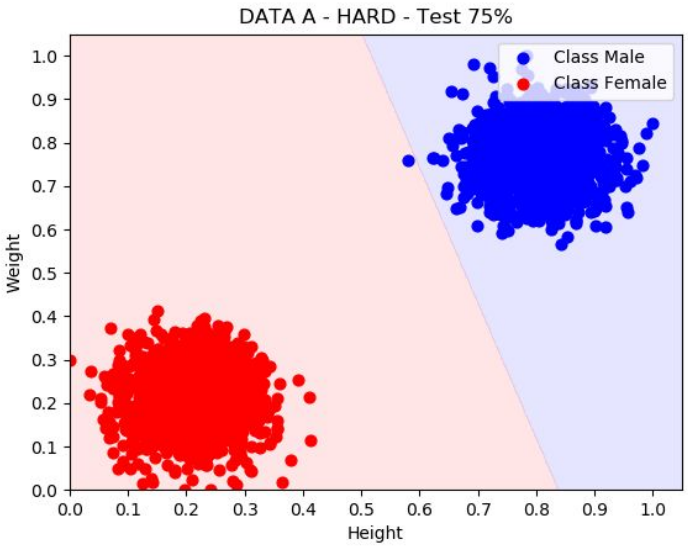




Confusion Matrix **TOTAL ERROR: 0.0**

Hard - Group A Testing 25%	Predicted F	Predicted M	Total	Recognition %
Actual F	500	0	500	100
Actual M	0	500	500	100
Total	500	500	1000	100

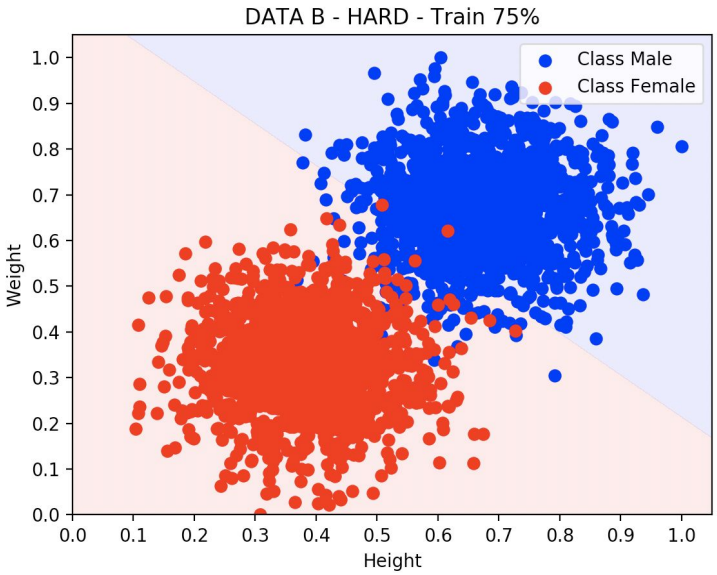


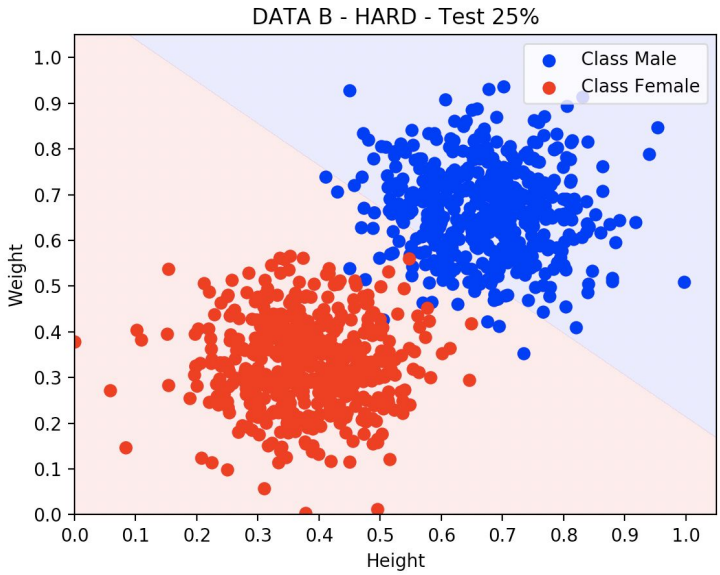


Confusion Matrix **TOTAL ERROR: 0.0**

Hard - Group A Testing 75%	Predicted F	Predicted M	Total	Recognition %
Actual F	1500	0	1500	100.0
Actual M	1	1499	1500	99.93
Total	1501	1499	3000	99.97

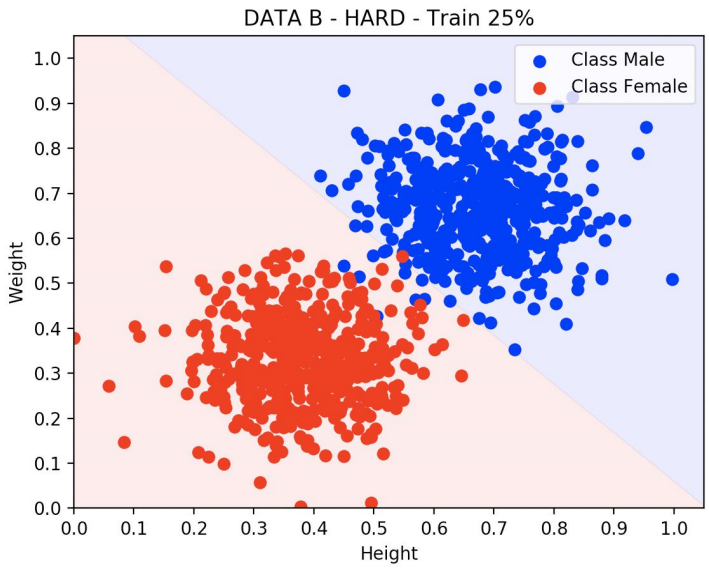
2.1 Group B Hard Activation Function

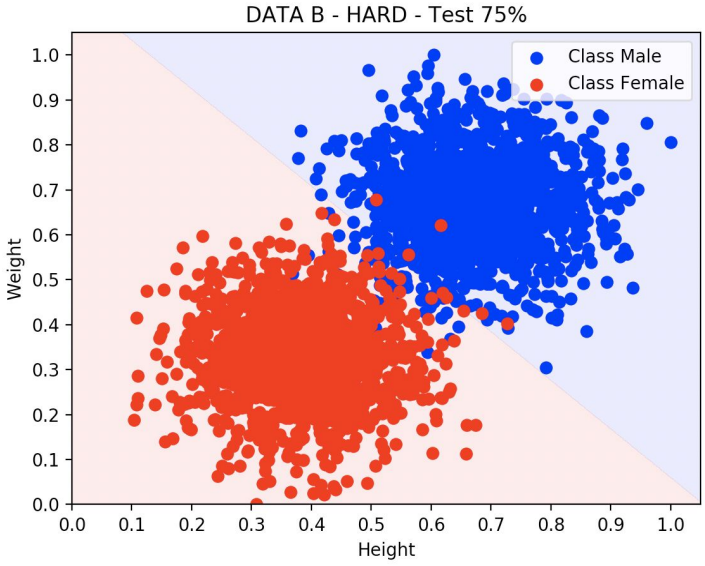




Confusion Matrix **Total Error: 94.0**

Hard - Group B Testing 25%	Predicted F	Predicted M	Total	Recognition %
Actual F	500	0	500	100.0
Actual M	57	443	500	88.6
Total	557	443	1000	94.3

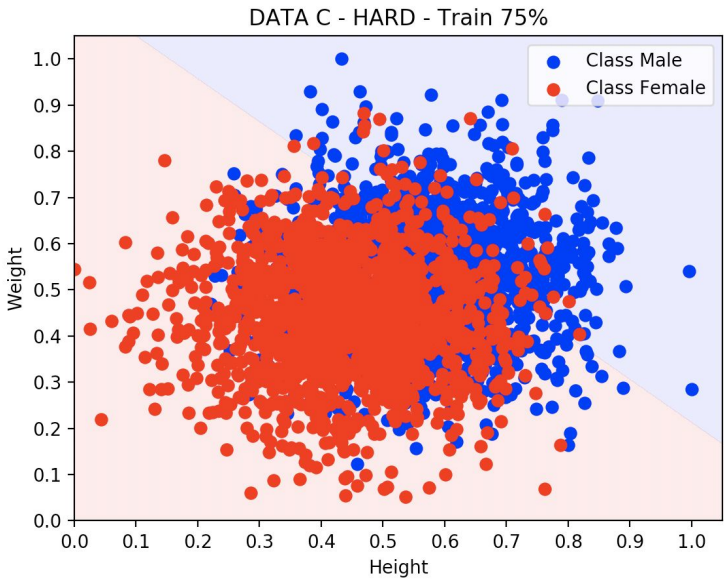


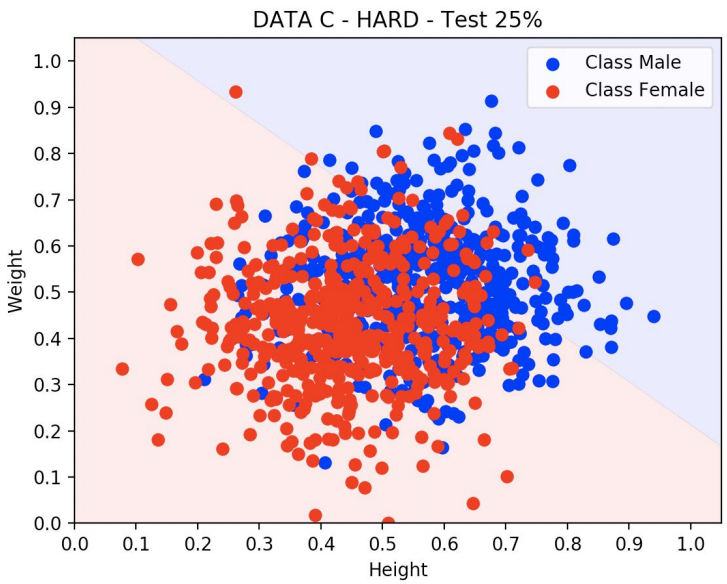


Confusion Matrix **TOTAL ERROR: 42.0**

Hard - Group B Testing 75%	Predicted F	Predicted M	Total	Recognition %
Actual F	1495	5	1500	99.67
Actual M	63	1437	1500	95.80
Total	1558	1442	3000	97.03

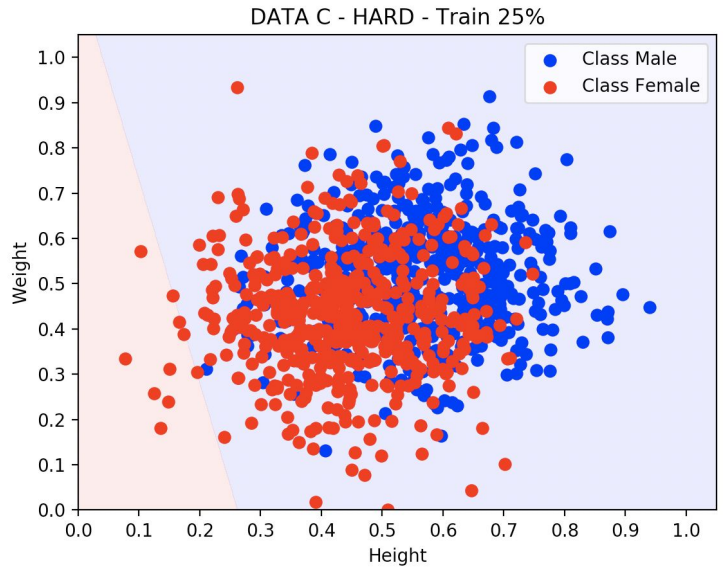
2.1 Group C Hard Activation Function

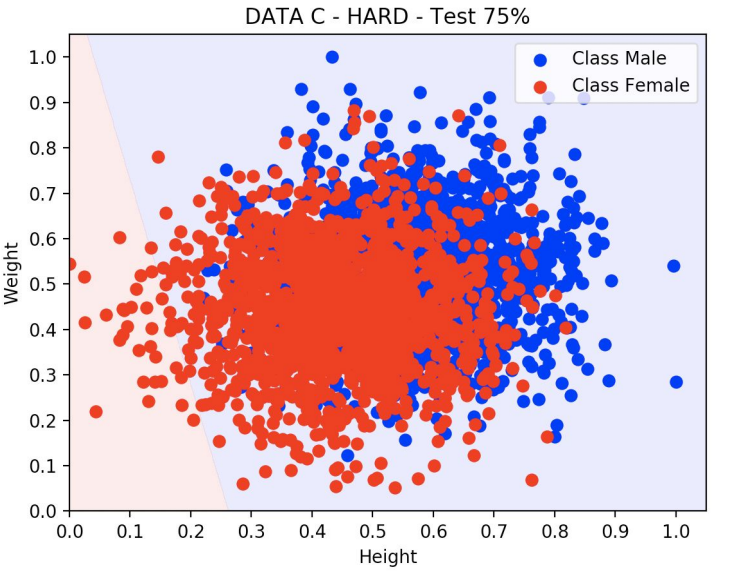




Confusion Matrix **TOTAL ERROR: 1092.0**

Hard - Group C Testing 25%	Predicted F	Predicted M	Total	Recognition %
Actual F	471	29	500	94.2
Actual M	344	156	500	31.2
Total	815	185	1000	62.7

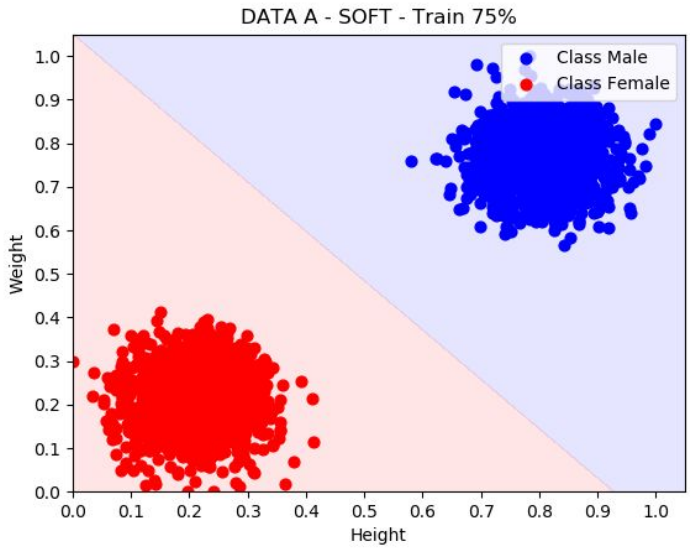


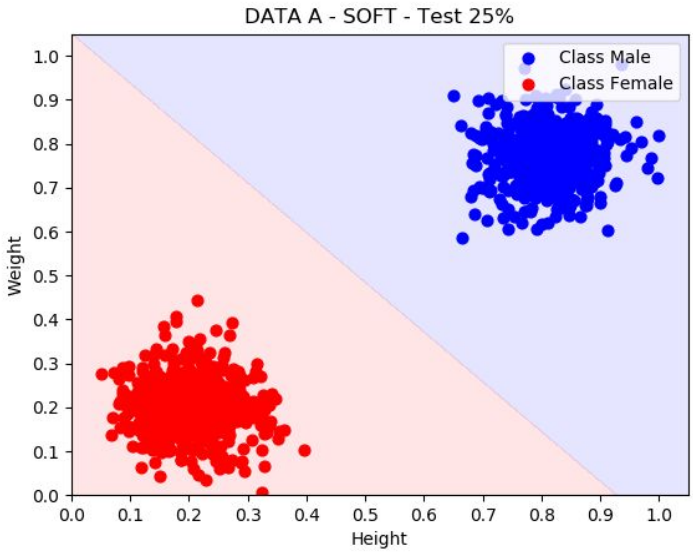


Confusion Matrix **TOTAL ERROR: 378.0**

Hard - Group C Testing 75%	Predicted F	Predicted M	Total	Recognition %
Actual F	31	1469	1500	2.07
Actual M	0	1500	1500	100.00
Total	31	2969	3000	51.03

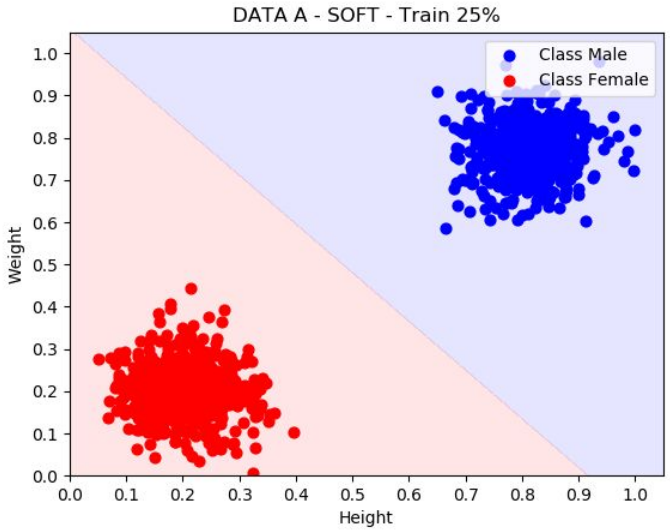
2.1 Group A Soft Activation Function

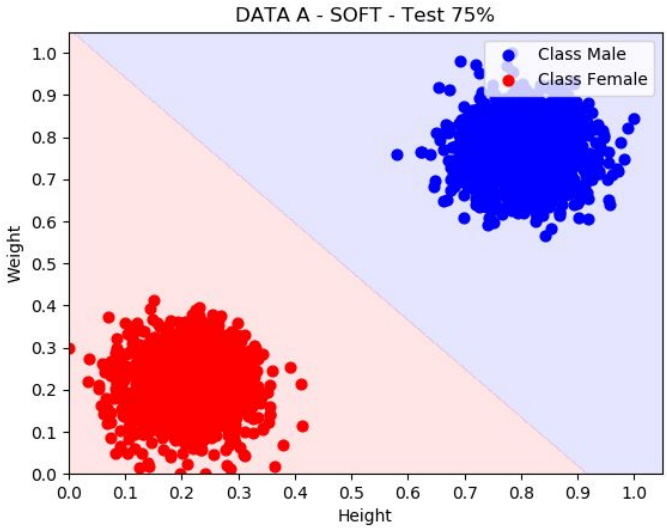




Confusion Matrix **TOTAL ERROR: 4.9959312620667086e-05**

Soft - Group A Testing 25%	Predicted F	Predicted M	Total	Recognition %
Actual F	500	0	500	100
Actual M	0	500	500	100
Total	500	500	1000	100

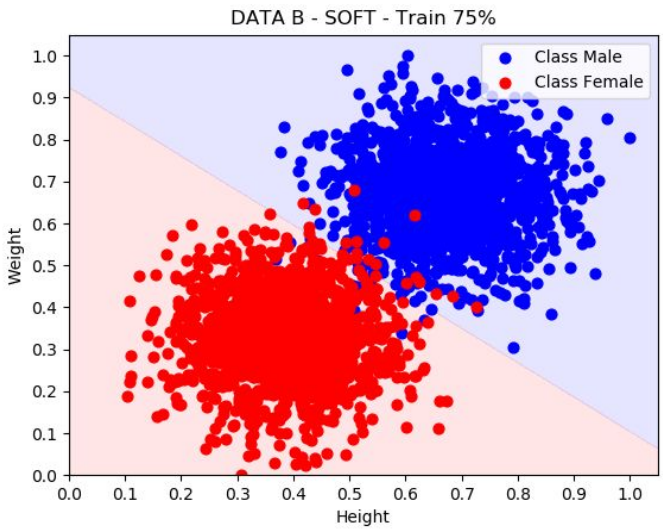


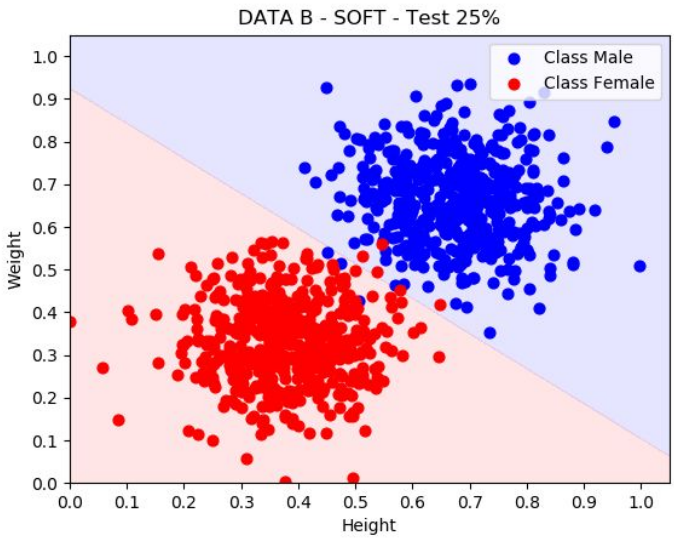


Confusion Matrix **TOTAL ERROR: 4.999980697335288e-05**

Soft - Group A Testing 75%	Predicted F	Predicted M	Total	Recognition %
Actual F	1500	0	1500	100
Actual M	0	1500	1500	100
Total	1500	1500	3000	100

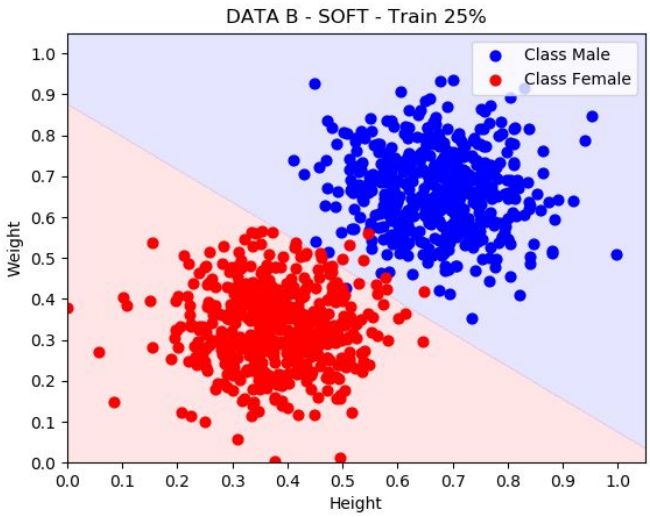
2.1 Group B Soft Activation Function

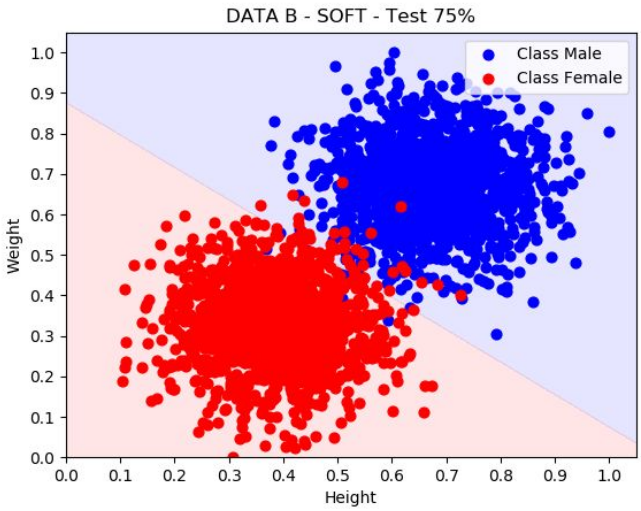




Confusion Matrix **TOTAL ERROR: 78.24029662169157**

Soft - Group B Testing 25%	Predicted F	Predicted M	Total	Recognition %
Actual F	496	4	500	99.2
Actual M	4	496	500	99.2
Total	500	500	1000	99.2

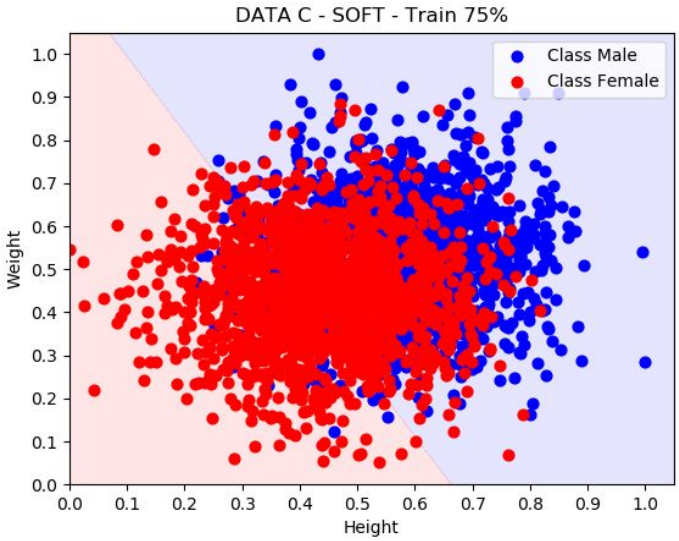


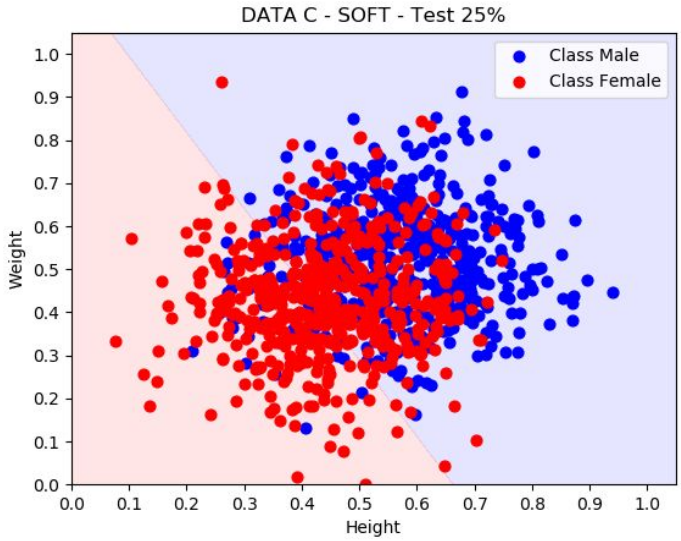


Confusion Matrix **TOTAL ERROR: 85.06004660468646**

Soft - Group B Testing 75%	Predicted F	Predicted M	Total	Recognition %
Actual F	1464	36	1500	97.60
Actual M	14	1486	1500	99.07
Total	1478	1522	3000	98.33

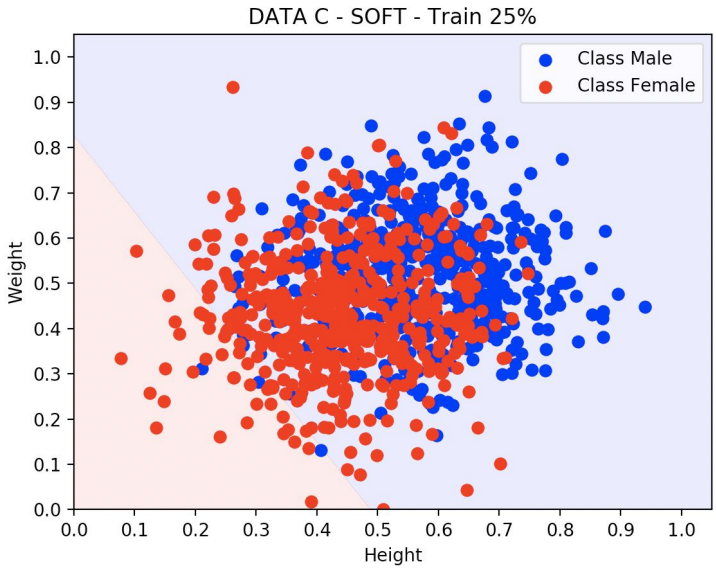
2.1 Group C Soft Activation Function

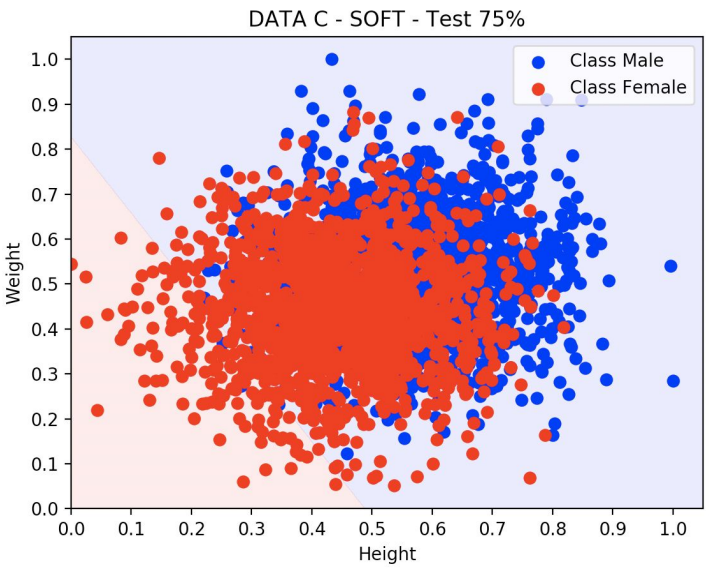




Confusion Matrix **TOTAL ERROR: 669.9315874219097**

Soft - Group C Testing 25%	Predicted F	Predicted M	Total	Recognition %
Actual F	241	259	500	48.20
Actual M	45	455	500	91.00
Total	286	714	1000	69.60





Confusion Matrix **TOTAL ERROR: 243.44301022517755**

Soft - Group C Test 75%	Predicted F	Predicted M	Total	Recognition %
Actual F	129	1371	1500	8.60
Actual M	4	1496	1500	99.73
Total	133	2867	3000	54.16

2.1. Compare 1. and 2. Are errors different and if so, why? What is the effect of different data set and effect of different training/testing distributions?

- **(Scenario A)** For Groups A and B, the hard activation function performed similarly regardless of the training and testing data was split. Generally speaking, when the model can train on 75% of the data, it usually classifies the remaining 25% more accurately, however for Group B, the datasets were distinct enough to where both the 25/75 and 75/25 splits had a high accuracy. This particular example also had a favorable neuron initialization, which wasn't frequently observed through our multiple experiments. In a heavily overlapped dataset lik Group C, the higher accuracy of the 75/25 split is clearly observed. Giving the model 75% of the data to train on gave it over 10% more accuracy than the 25/75 split.
- **(Scenario B)** For the soft activation function, the differences in error were similar to those observed for scenario A. The effects of different data sets and training/testing distributions affected it as well. In general, the model classified the data set more accurately with a 75/25 split, since it had more data to train on and modify its weights more accurately. The difference in data overlap for each dataset also affected it, as more overlapped data makes it more difficult to classify data.

2.1b.When would you use option 1 and option 2 above?

- We'd use 75/25 over 25/75 in most scenarios, especially if working with a hard activation function, since we want to ensure the line is as well placed as possible. We'd use 25/75 in situations where we don't want to risk overfitting the model, to allow for it to accurately classify new data coming in. Training on 25% could potentially be used to find bad data, suppose we find that a particular sector of data has caused our function to result in an unfavorable bisecting line, we could reasonably conclude that some issue exists within that data.

2.2b. Compare and discuss results when hard activation was used vs. when soft activation was used.

- The soft activation function, consistently had higher accuracy than the hard activation function in nearly every scenario. While the hard function's accuracy was not too far from the soft, it's clear that in certain scenarios (such as when it trained on 25% of the data for Group A), hard would perform significantly worse if new data was introduced, as the line it generated was right at the edge of one cluster. For the same scenario, the soft function not only accurately classified the data, but produced a line that would likely remain accurate if new data points was introduced.