CMSC 409: Artificial Intelligence
Project No. 2
Due Thursday, October 3, 2023, noon

Student Certification:
Team Member 1:
Print Name: Stephen Liu Date: 10/3
I have contributed by doing the following:_____

Signed: Pratham Choksi

Team Member 2:
Print Name: Sajesh Sahoo Date: 10/3
I have contributed by doing the following: Helped code the activation functions, as well as the graph method, and the test function helped with the documentation.

Signed: Sajesh Sahoo

Team Member 3:
Print Name: Pratham Choksi Date: 10/3
I have contributed by doing the following: _____

Signed: Stephen Liu

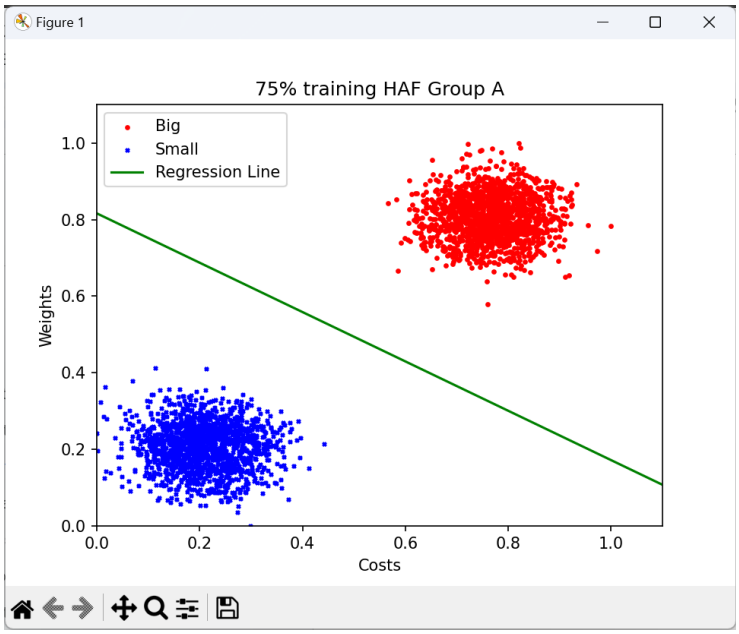
HAF

Dataset A

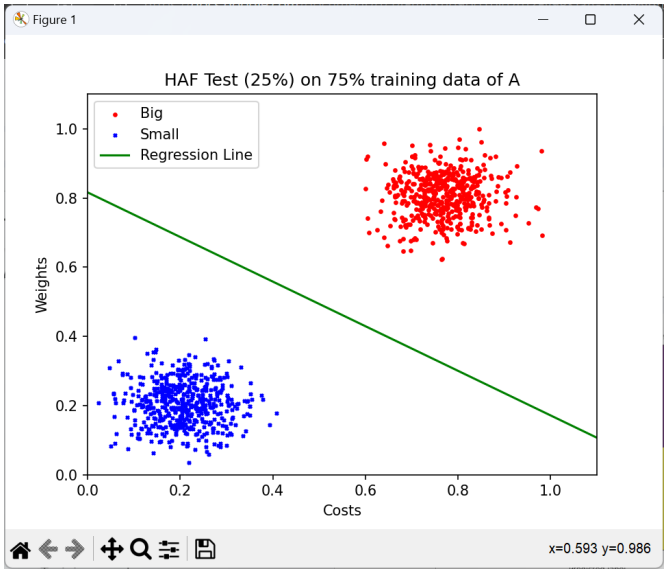
75% training, 25% testing

Total error: 7

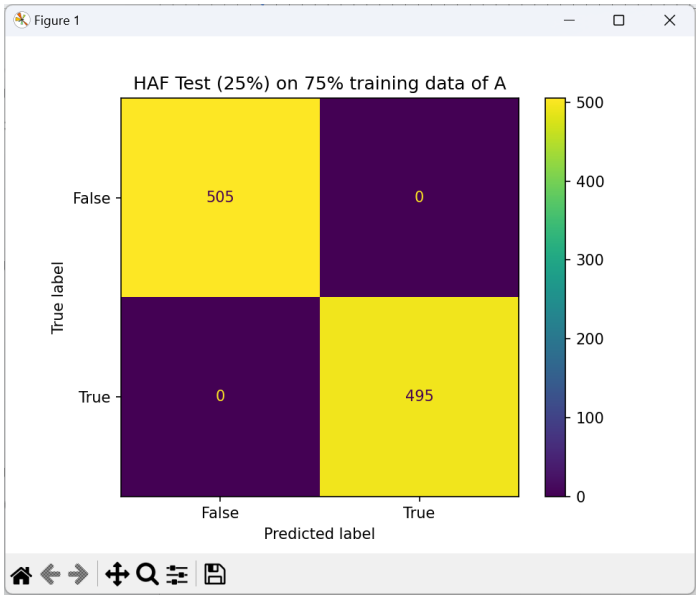
Training graph



Testing graph



Confusion matrix



Analysis

Accuracy: 100%
Error rate: 0%
TPR: 100%
TNR: 100%
FPR: 0%
FNR: 0%

From project 1:

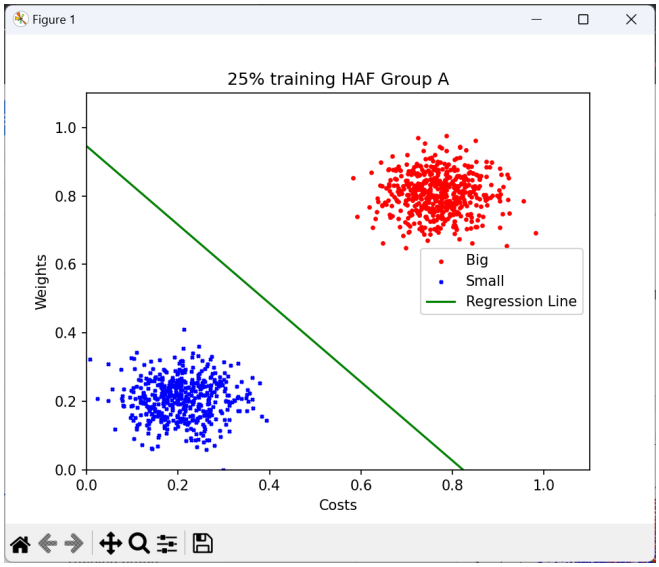
True Positive Rate: 1
True Negative Rate: 1
False Positive Rate: 0
False Negative Rate: 0

Rates are the same.

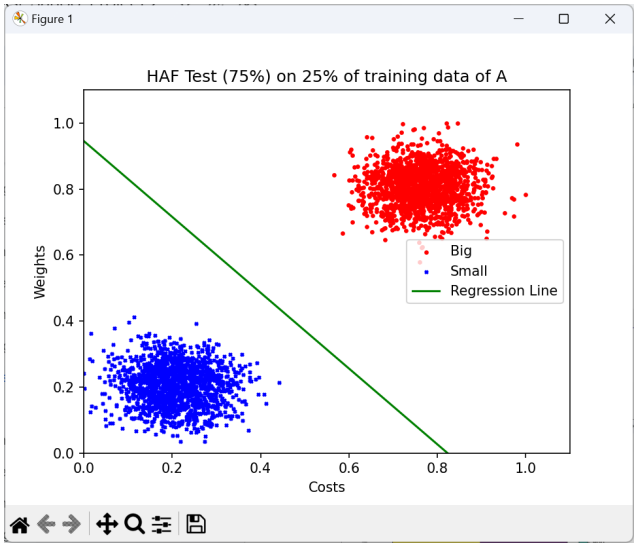
25% training, 75% testing

Total error: 3

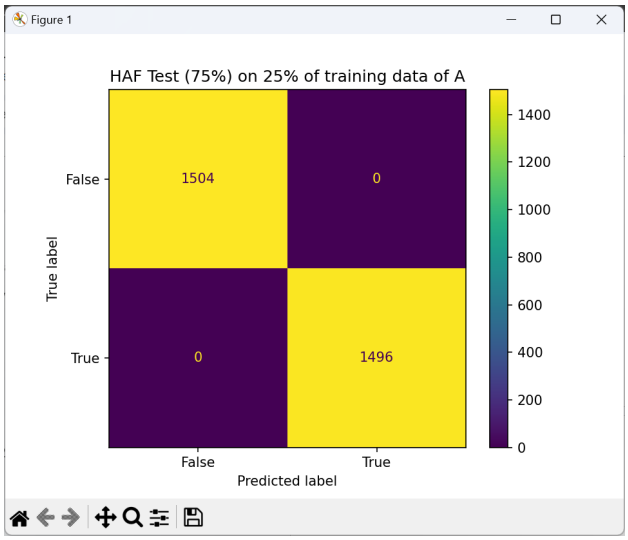
Training graph



Testing graph



Confusion matrix



Analysis

Accuracy: 100%
Error rate: 0%
TPR: 100%
TNR: 100%
FPR: 0%
FNR: 0%

From project 1:
True Positive Rate: 1
True Negative Rate: 1

False Positive Rate: 0
False Negative Rate: 0

Rates are the same.

Analysis

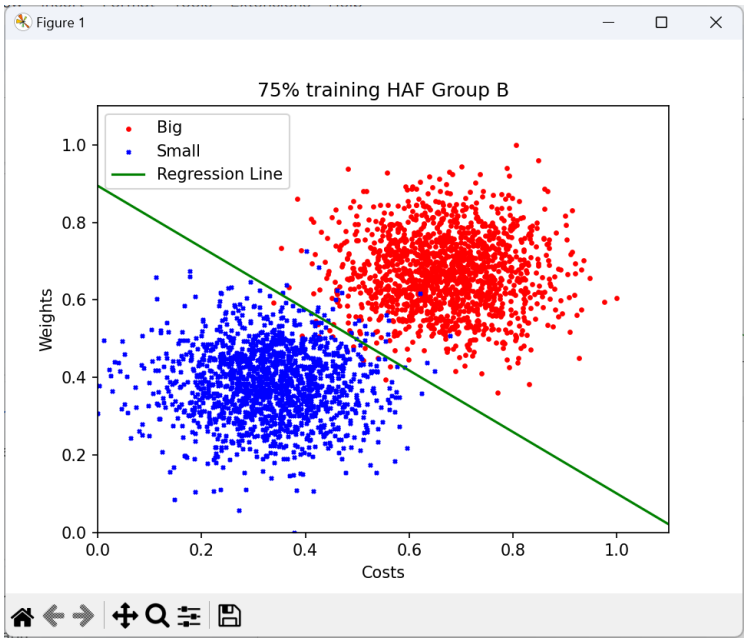
Total error was almost double when training dataset was 75% of original dataset, likely because bigger training dataset meant more patterns to add error to the total.

Dataset B

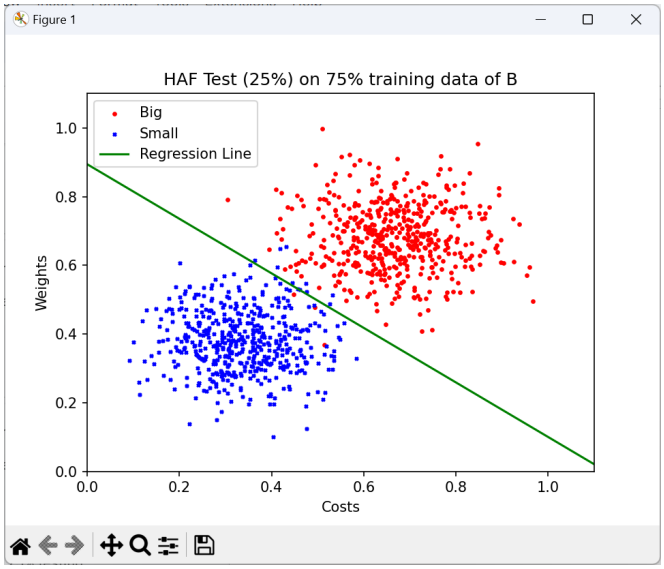
75% training, 25% testing

Total error: 99

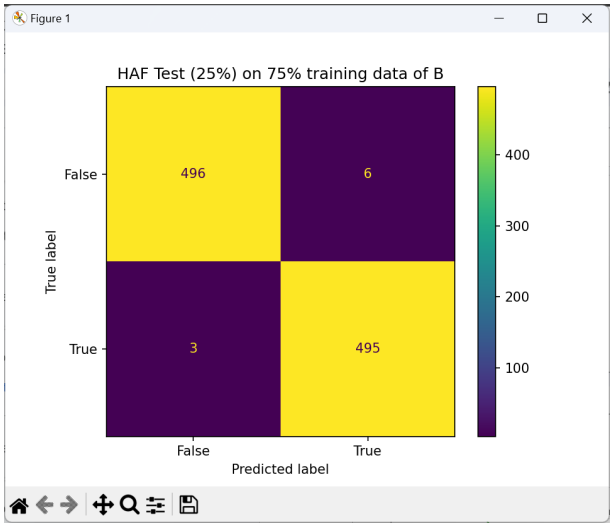
Training graph



Testing graph



Confusion matrix



Analysis

Accuracy: 99.1%
Error rate: 0.9%
TPR: 99.4%
TNR: 98.8%
FPR: 1.2%
FNR: 0.6%

From Project 1:
True Positive Rate: .991
True Negative Rate: .9845

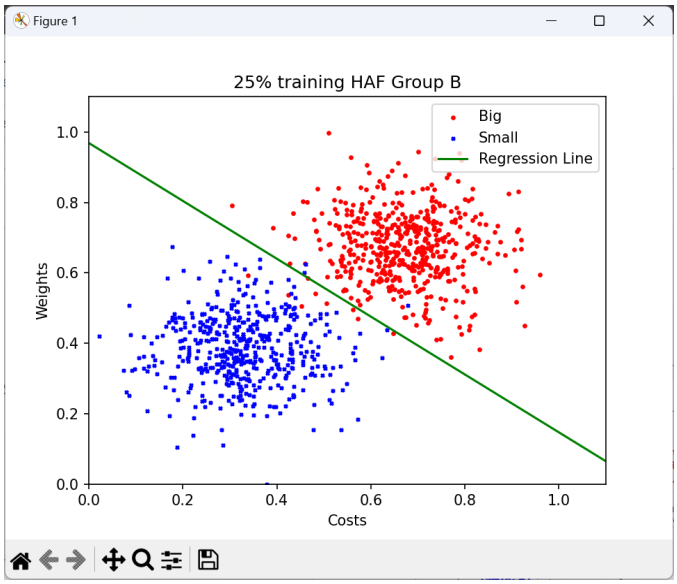
False Positive Rate: .0155
False Negative Rate: .009

TPR and TNR are slightly higher in project 2 than project 1.

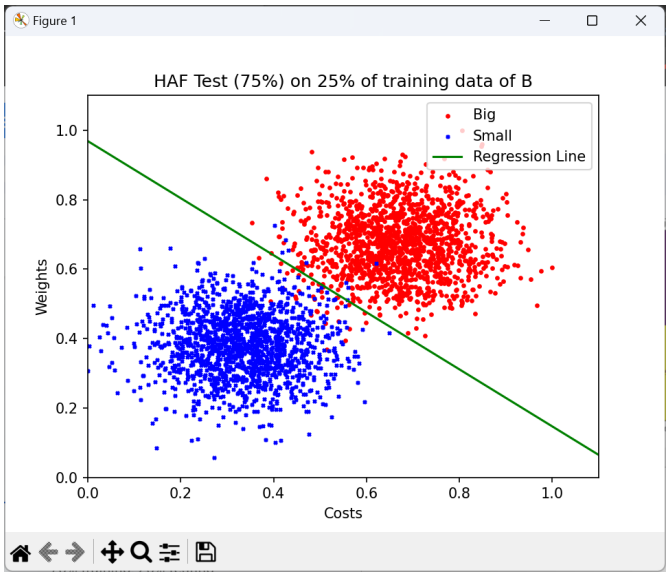
25% training, 75% testing

Total error: 41

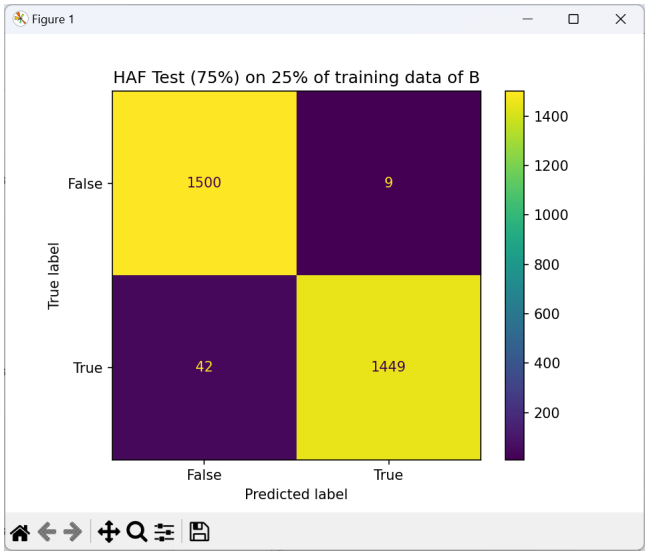
Training graph



Testing graph



Confusion matrix



Analysis

Accuracy: 98.3%
Error rate: 1.7%
TPR: 97.2%
TNR: 99.4%
FPR: 0.6%
FNR: 2.8%

From Project 1:
True Positive Rate: .991
True Negative Rate: .9845
False Positive Rate: .0155
False Negative Rate: .009

TN and FN rates are slightly higher in project 2 than project 1.

Analysis

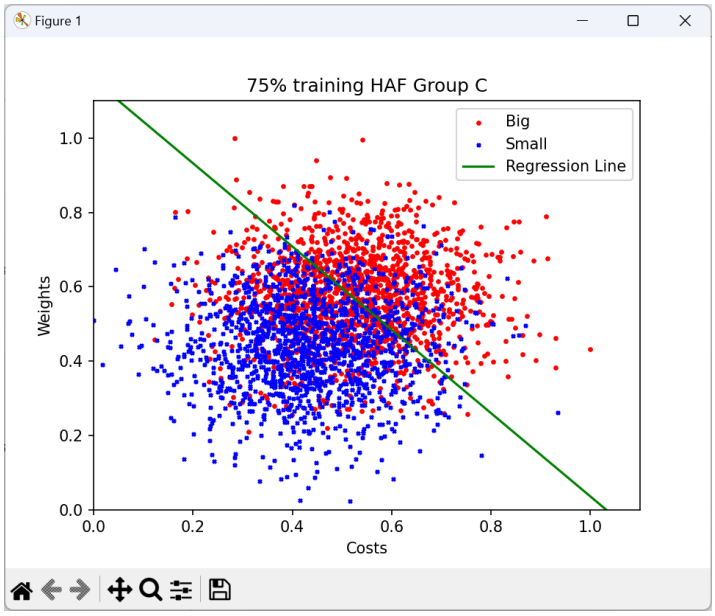
Total error was almost double when training dataset was 75% of original dataset, likely because bigger training dataset meant more patterns to add error to the total.

Dataset C

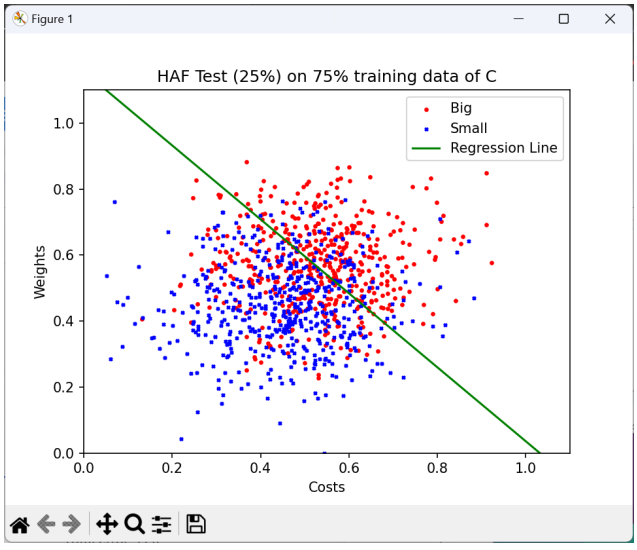
75% training, 25% testing

Total error: 803

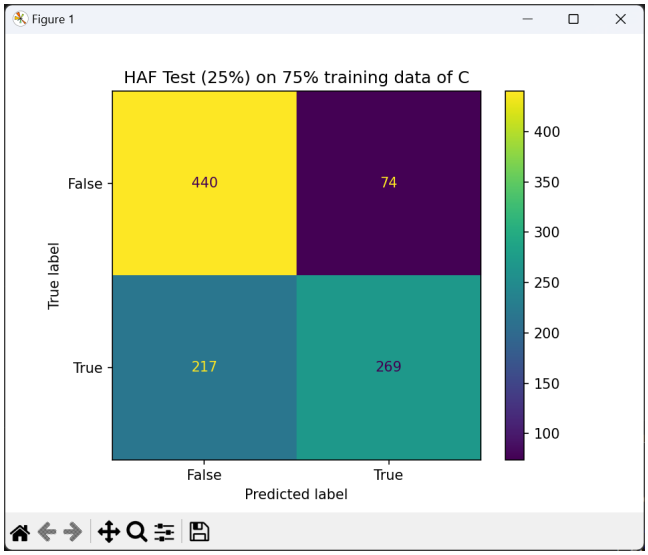
Training graph



Testing graph



Confusion matrix



Analysis

Accuracy: 70.9%
Error rate: 29.1%
TPR: 55.4%
TNR: 85.6%
FPR: 14.4%
FNR: 44.6%

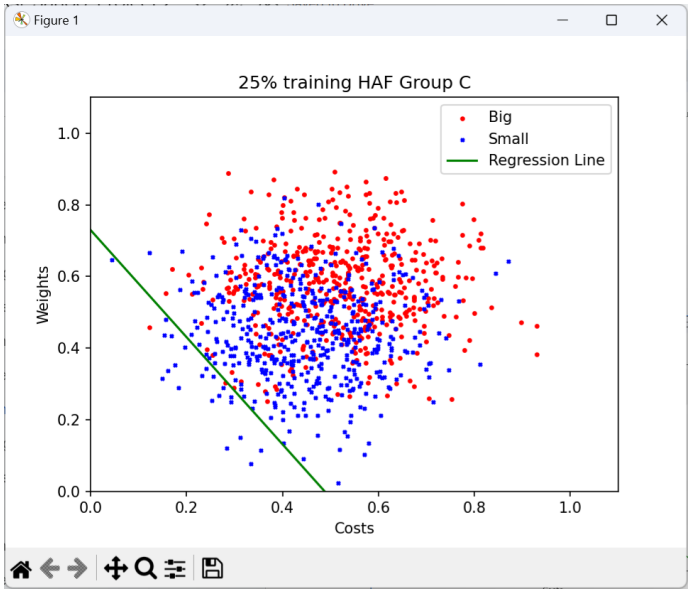
From project 1:
True Positive Rate: .5915
True Negative Rate: .826
False Positive Rate: .174
False Negative Rate: .4085

TN and FN are slightly higher in project 2 than project 1.

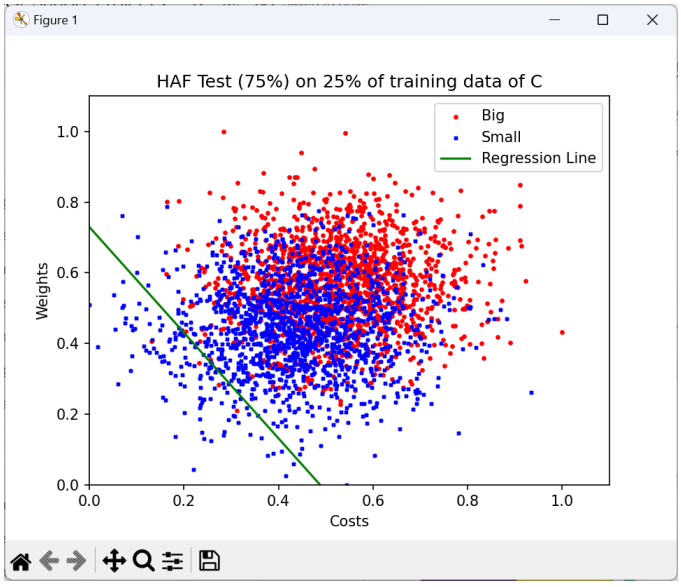
25% training, 75% testing

Total error: 726

Training graph



Testing graph



Confusion matrix

Analysis

Accuracy: 52.8%
Error rate: 47.2%
TPR: 99.9%
TNR: 6.1%
FPR: 93.9%
FNR: 0.01%

From project 1:
True Positive Rate: .5915
True Negative Rate: .826
False Positive Rate: .174
False Negative Rate: .4085

TP and FP rates are much higher in project 2 than project 1.

Analysis

Total error was somewhat higher when training dataset was 75% of original dataset, likely because bigger training dataset meant more patterns to add error to the total.

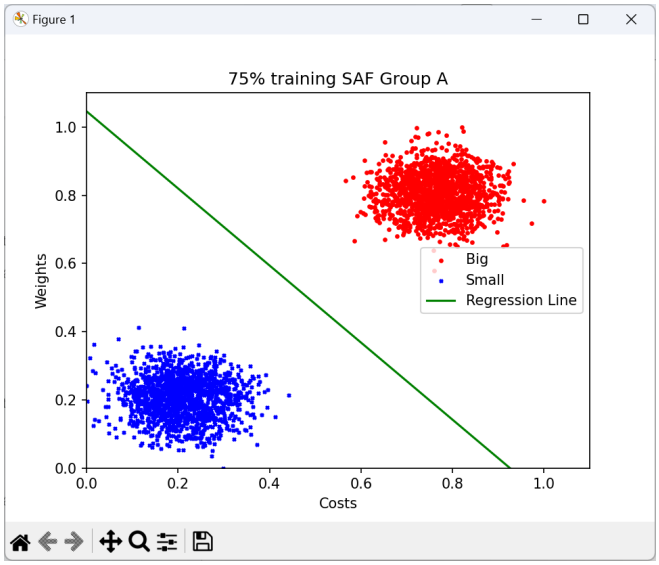
SAF

Dataset A

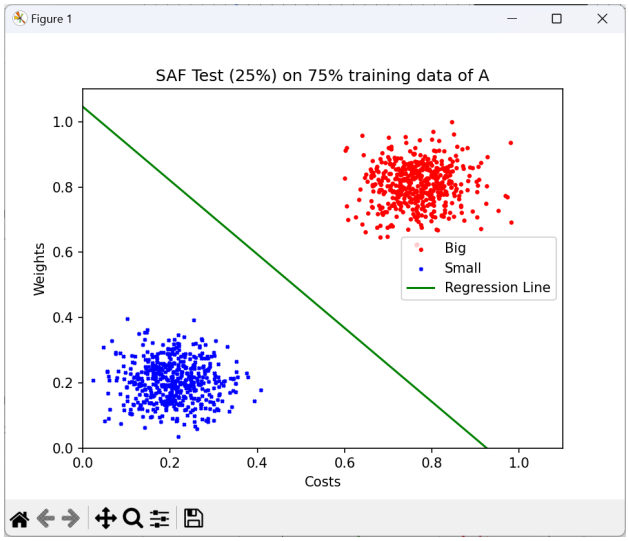
75% training, 25% testing

Total error: 92.535493

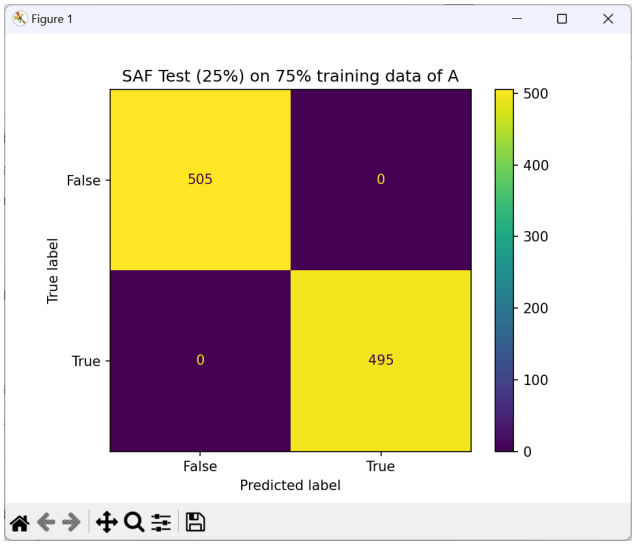
Training graph



Testing graph



Confusion matrix



Analysis

Accuracy: 100%
Error rate: 0%
TPR: 100%
TNR: 100%
FPR: 0%
FNR: 0%

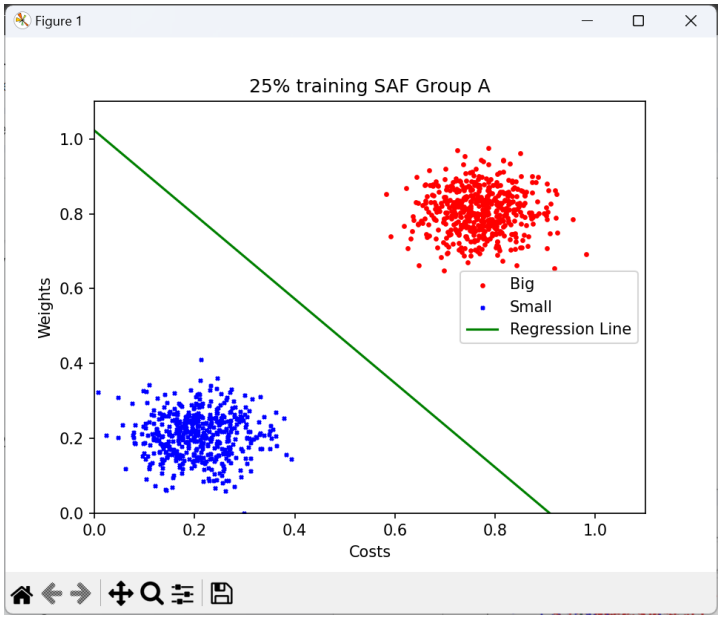
From project 1:
Accuracy:1
Error:0
True Positive Rate: 1
True Negative Rate: 1
False Positive Rate: 0
False Negative Rate: 0

Rates are the same.

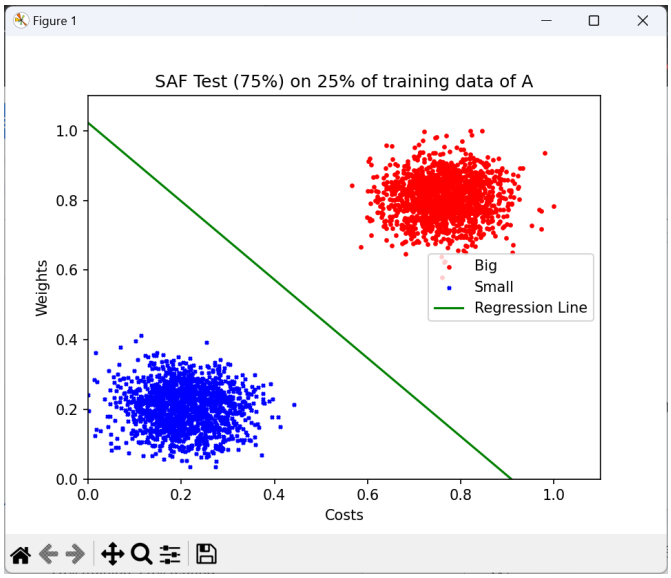
25% training, 75% testing

Total error: 44.20363975

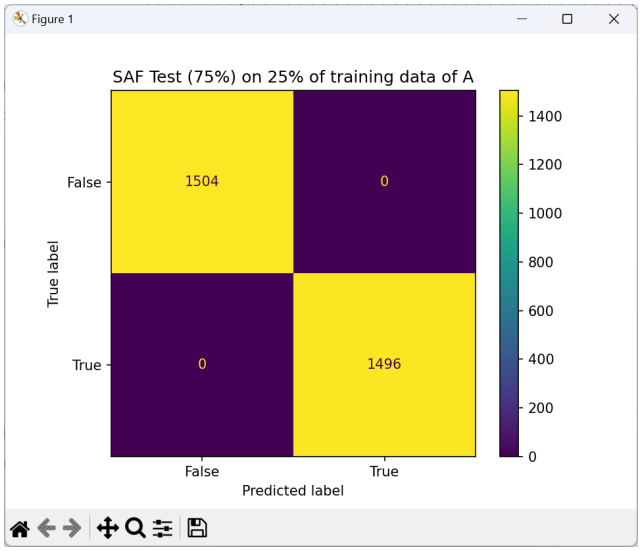
Training graph



Testing graph



Confusion matrix



Analysis

Accuracy: 100%
Error rate: 0%
TPR: 100%
TNR: 100%
FPR: 0%
FNR: 0%

From project 1:
Accuracy:1
Error:0
True Positive Rate: 1
True Negative Rate: 1
False Positive Rate: 0
False Negative Rate: 0

Rates are the same.

Analysis

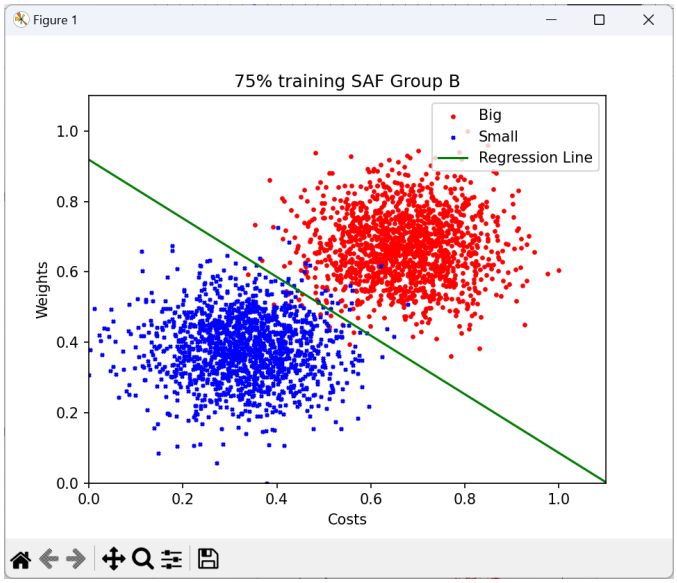
Total error was almost double when training dataset was 75% of original dataset, likely because bigger training dataset meant more patterns to add error to the total.

Dataset B

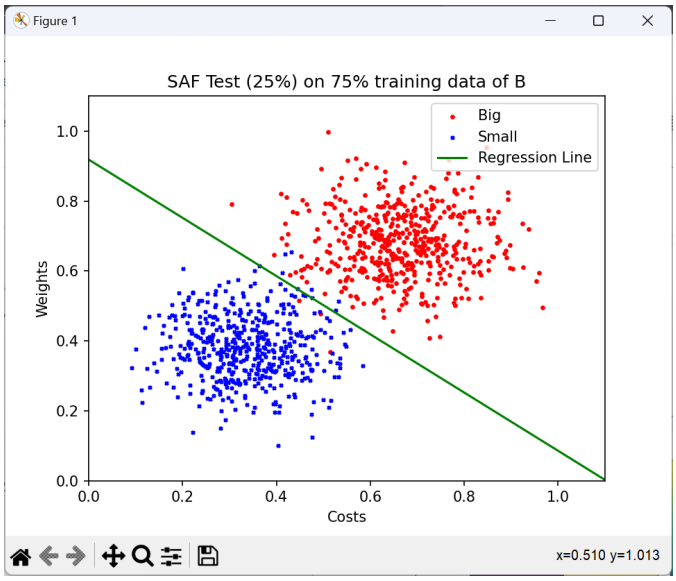
75% training, 25% testing

Total error: 306.86210388

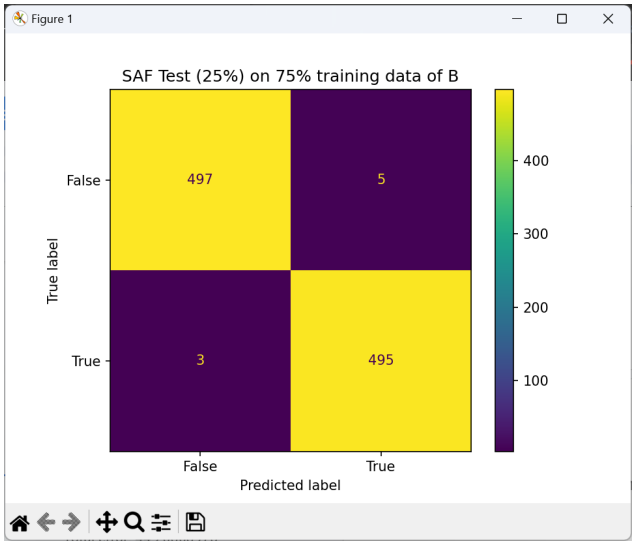
Training graph



Testing graph



Confusion matrix



Analysis

Accuracy: 99.2%
Error rate: 0.8%
TPR: 99.4%
TNR: 99%
FPR: 1%
FNR: 0.6%

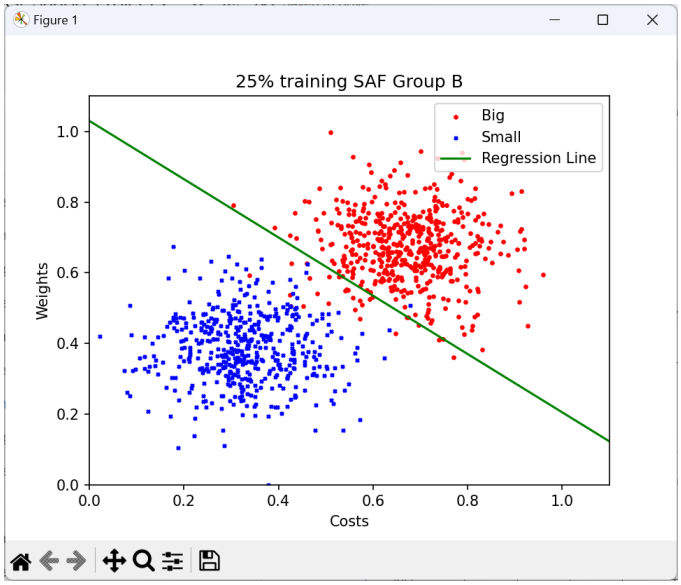
From Project 1:
True Positive Rate: .991
True Negative Rate: .9845
False Positive Rate: .0155
False Negative Rate: .009

TP and TN rates are slightly higher in Project 2 than Project 1.

25% training, 75% testing

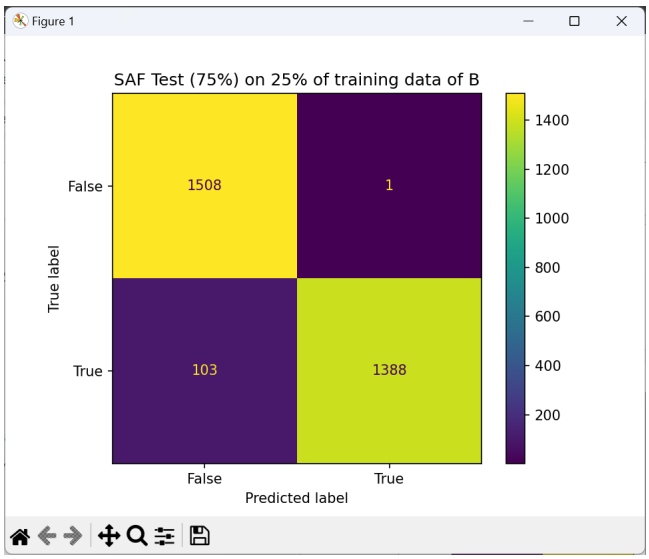
Total error: 81.905208

Training graph



Testing graph

Confusion matrix



Analysis

Accuracy: 96.5%
Error rate: 3.5%
TPR: 93.1%
TNR: 99.9%
FPR: 0.1%
FNR: 6.9%

From Project 1:
True Positive Rate: .991
True Negative Rate: .9845
False Positive Rate: .0155
False Negative Rate: .009

TN and FN rates are somewhat higher in project 2 than project 1.

Analysis

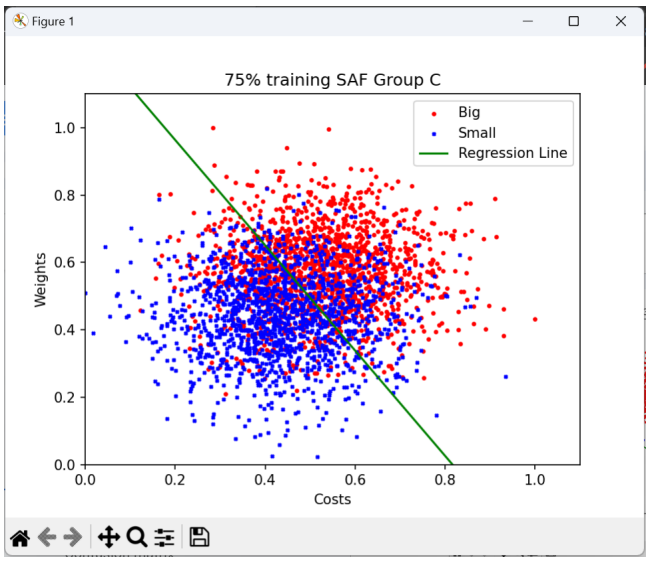
Total error was much higher when training dataset was 75% of original dataset, likely because bigger training dataset meant more patterns to add error to the total.

Dataset C

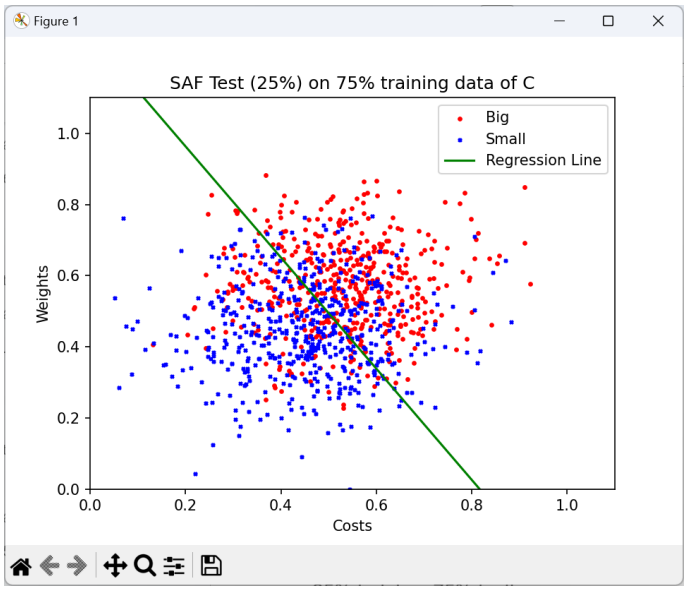
75% training, 25% testing

Total error: 895.4816300

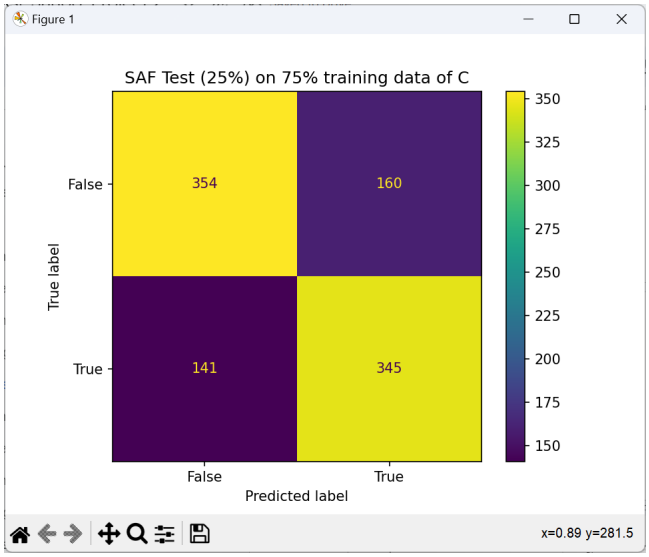
Training graph



Testing graph



Confusion matrix



Analysis

Accuracy: 69.9%
Error rate: 30.1%
TPR: 71%
TNR: 68.9%
FPR: 31.1%
FNR: 29%

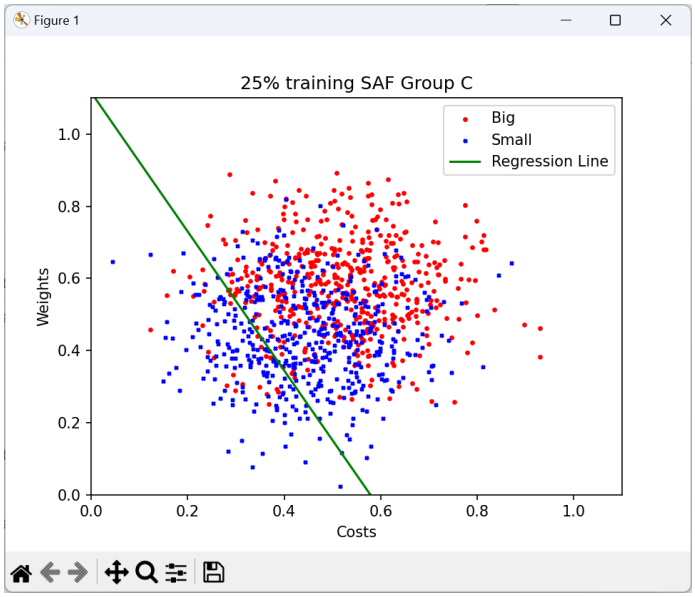
From project 1:
True Positive Rate: .5915
True Negative Rate: .826
False Positive Rate: .174
False Negative Rate: .4085

TP and FP rates are significantly higher in project 2 than project 1.

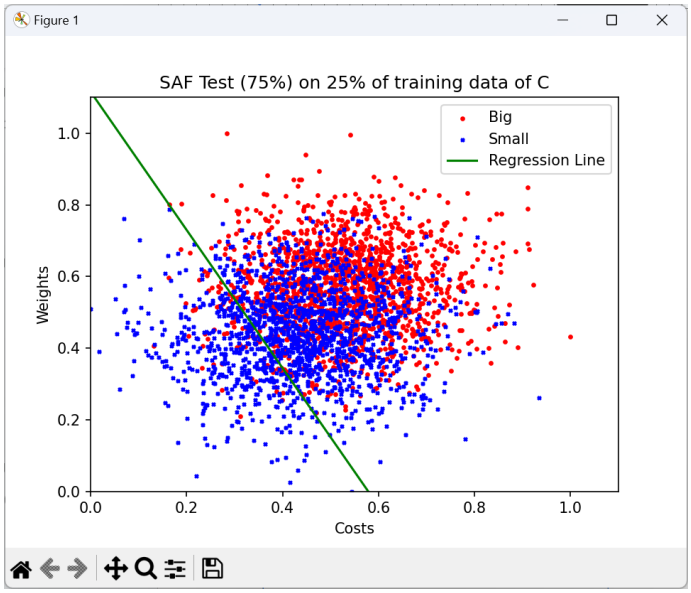
25% training, 75% testing

Total error: 718.2794685

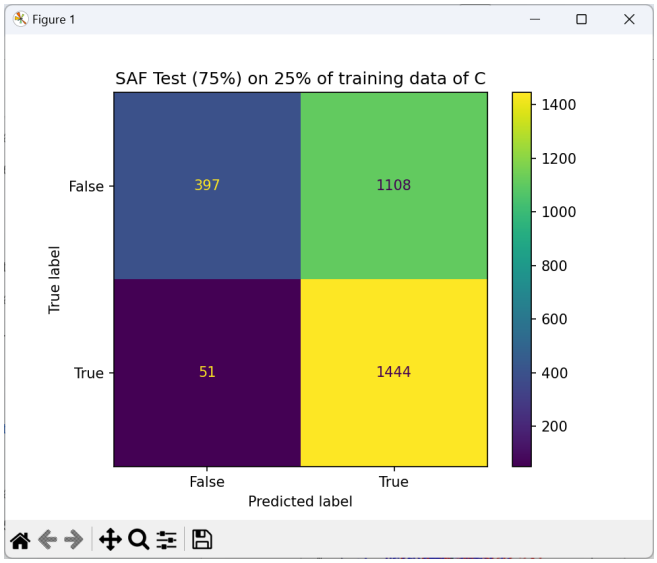
Training graph



Testing graph



Confusion matrix



Analysis

Accuracy: 61.5%
Error rate: 38.5%
TPR: 96.6%
TNR: 26.4%
FPR: 73.6%
FNR: 3.4%

From project 1:
True Positive Rate: .5915
True Negative Rate: .826
False Positive Rate: .174
False Negative Rate: .4085

TP and FP rates are significantly higher in project 2 than project 1.

Analysis

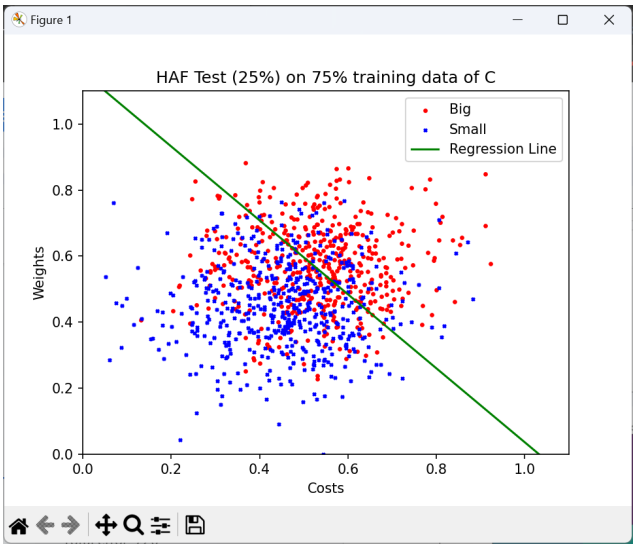
Total error was somewhat higher when training dataset was 75% of original dataset, likely because bigger training dataset meant more patterns to add error to the total.

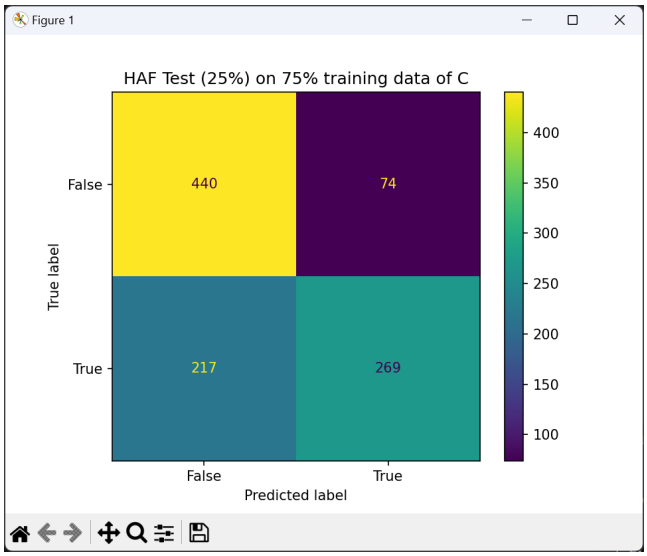
Overall analysis

For every dataset, larger training datasets resulted in larger total errors but smaller error rates (higher accuracy). For both HAF and SAF, accuracy was highest for dataset A. Dataset B had the next highest accuracy, followed by dataset C. I'd probably go with a larger proportion training dataset if the initial dataset was small or skewed to one class, so that the neuron learns from a wide variety of points. However, if the initial dataset was very large, it'd probably be better to select a smaller proportion training dataset to speed up the training process.

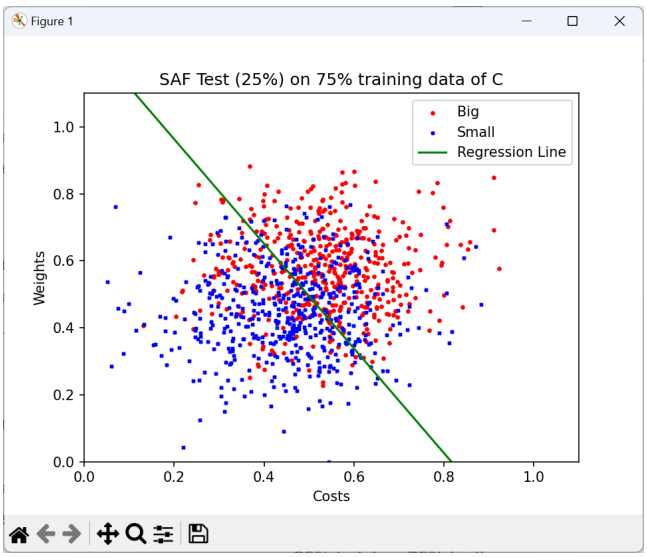
Between identical datasets and training dataset proportions, HAF had slightly higher accuracy except for 25% training on dataset C, where SAF had the higher accuracy. It also had larger total error.

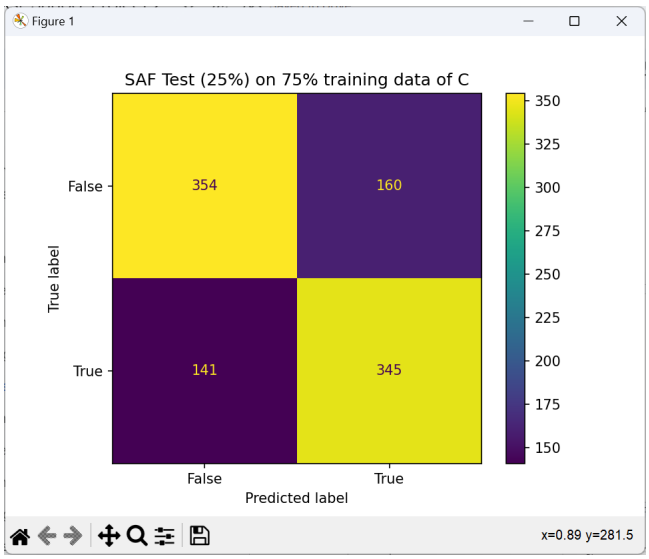
HAF, Dataset C, 25% testing:





SAF, Dataset C, 25% testing





Extra credit:

I think it's best to have the training dataset cover all types of classes. Even if it can't be evenly distributed, it should include less common classes more often and more common classes less often. This is so that the neuron is trained to recognize all classes. This is very important, since often the most likely outcome isn't the most important one. For example, if we were developing a machine learning algorithm to identify a rare disease diagnosis from medical readings, it's likely that most of the data points it encounters will be from people who don't have it. But if it doesn't recognize when the rare disease is present, it's worse than useless and can cause major harm to patients.