\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

CAB320 ASSIGNMENT 2: MACHINE LEARNING

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

TEAM MEMBERS:

10155856 : Mackenzie Wilson

10157182 : Nicole Barritt

Training set: 80.0% of data, 455 samples, 170 positive (37.4%)

Test set: 20.0% of data, 114 samples, 42 positive (36.8%)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* NAIVE BAYES CLASSIFIER

Working...

- - - VALIDATION REPORT - - -

TRAINING VALIDATION

accuracy: 0.39 0.39

precision: 0.36 0.36

recall: 0.84 0.83

f1: 0.51 0.51

roc\_auc: 0.49 0.48

Mean fit time: 0.001225 seconds

Mean score time: 0.003675 seconds

Cross-validation, optimization, and fitting took 0.030858 seconds total.

- - - TRAINING REPORT - - -

Confusion Matrix:

PREDICTED: Malignant Benign All

ACTUAL:

Malignant (TP) 142 (FN) 28 170

Benign (FP) 248 (TN) 37 285

All 390 65 455

Classification Report:

precision recall f1-score support

Malignant 0.36 0.84 0.51 170

Benign 0.57 0.13 0.21 285

avg / total 0.49 0.39 0.32 455

Overall Metrics:

accuracy: 0.39

roc\_auc: 0.48

precision: 0.36

recall: 0.84

f1: 0.51

lgrthmc loss: 20.95

mse: 0.61

variance: -0.59

Prediction on training set took 0.000413 seconds.

- - - TEST REPORT - - -

Confusion Matrix:

PREDICTED: Malignant Benign All

ACTUAL:

Malignant (TP) 35 (FN) 7 42

Benign (FP) 66 (TN) 6 72

All 101 13 114

Classification Report:

precision recall f1-score support

Malignant 0.35 0.83 0.49 42

Benign 0.46 0.08 0.14 72

avg / total 0.42 0.36 0.27 114

Overall Metrics:

accuracy: 0.36

roc\_auc: 0.46

precision: 0.35

recall: 0.83

f1: 0.49

lgrthmc loss: 22.12

mse: 0.64

variance: -0.60

Prediction on test set took 0.000145 seconds.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* DECISION TREE CLASSIFIER

Hyperparameter: max\_depth

Values Tested: 100 values from 1 to 100

Working...

- - - VALIDATION REPORT - - -

Based on validation roc\_auc scores, the best value for hyperparameter 'max\_depth' is:

4

TRAINING VALIDATION

accuracy: 0.99 0.94

roc\_auc: 1.00 0.94

precision: 0.99 0.90

recall: 0.99 0.94

f1: 0.99 0.92

Mean fit time: 0.005057 seconds

Mean score time: 0.004246 seconds



Zoomed in graph:



Cross-validation, optimization, and fitting took 3.832448 seconds total.

- - - TRAINING REPORT - - -

Confusion Matrix:

PREDICTED: Malignant Benign All

ACTUAL:

Malignant (TP) 169 (FN) 1 170

Benign (FP) 0 (TN) 285 285

All 169 286 455

Classification Report:

precision recall f1-score support

Malignant 1.00 0.99 1.00 170

Benign 1.00 1.00 1.00 285

avg / total 1.00 1.00 1.00 455

Overall Metrics:

accuracy: 1.00

roc\_auc: 1.00

precision: 1.00

recall: 0.99

f1: 1.00

lgrthmc loss: 0.08

mse: 0.00

variance: 0.99

Prediction on training set took 0.001014 seconds.

- - - TEST REPORT - - -

Confusion Matrix:

PREDICTED: Malignant Benign All

ACTUAL:

Malignant (TP) 37 (FN) 5 42

Benign (FP) 1 (TN) 71 72

All 38 76 114

Classification Report:

precision recall f1-score support

Malignant 0.97 0.88 0.93 42

Benign 0.93 0.99 0.96 72

avg / total 0.95 0.95 0.95 114

Overall Metrics:

accuracy: 0.95

roc\_auc: 0.93

precision: 0.97

recall: 0.88

f1: 0.93

lgrthmc loss: 1.82

mse: 0.05

variance: 0.78

Prediction on test set took 0.000381 seconds.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* NEAREST NEIGHBOR CLASSIFIER

Hyperparameter: n\_neighbors

Values Tested: 200 values from 1 to 200

Working...

- - - VALIDATION REPORT - - -

Based on validation roc\_auc scores, the best value for hyperparameter 'n\_neighbors' is:

2

TRAINING VALIDATION

accuracy: 0.86 0.76

roc\_auc: 0.97 0.76

precision: 1.00 0.90

recall: 0.61 0.40

f1: 0.76 0.55

Mean fit time: 0.000706 seconds

Mean score time: 0.009211 seconds



Zoomed in graph: 

Cross-validation, optimization, and fitting took 39.484194 seconds total.

- - - TRAINING REPORT - - -

Confusion Matrix:

PREDICTED: Malignant Benign All

ACTUAL:

Malignant (TP) 116 (FN) 54 170

Benign (FP) 0 (TN) 285 285

All 116 339 455

Classification Report:

precision recall f1-score support

Malignant 1.00 0.68 0.81 170

Benign 0.84 1.00 0.91 285

avg / total 0.90 0.88 0.88 455

Overall Metrics:

accuracy: 0.88

roc\_auc: 0.84

precision: 1.00

recall: 0.68

f1: 0.81

lgrthmc loss: 4.10

mse: 0.12

variance: 0.55

Prediction on training set took 0.004244 seconds.

- - - TEST REPORT - - -

Confusion Matrix:

PREDICTED: Malignant Benign All

ACTUAL:

Malignant (TP) 17 (FN) 25 42

Benign (FP) 1 (TN) 71 72

All 18 96 114

Classification Report:

precision recall f1-score support

Malignant 0.94 0.40 0.57 42

Benign 0.74 0.99 0.85 72

avg / total 0.82 0.77 0.74 114

Overall Metrics:

accuracy: 0.77

roc\_auc: 0.70

precision: 0.94

recall: 0.40

f1: 0.57

lgrthmc loss: 7.88

mse: 0.23

variance: 0.21

Prediction on test set took 0.001551 seconds.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* SUPPORT VECTOR MACHINE CLASSIFIER

Hyperparameter: gamma

Values Tested: 100 values from 1e-10 to 10.0

Working...

- - - VALIDATION REPORT - - -

Based on validation roc\_auc scores, the best value for hyperparameter 'gamma' is:

1e-06

TRAINING VALIDATION

accuracy: 0.96 0.76

roc\_auc: 0.99 0.87

precision: 0.98 0.83

recall: 0.90 0.46

f1: 0.94 0.59

Mean fit time: 0.005973 seconds

Mean score time: 0.014064 seconds



Zoomed in graph:



Cross-validation, optimization, and fitting took 13.931500 seconds total.

- - - TRAINING REPORT - - -

Confusion Matrix:

PREDICTED: Malignant Benign All

ACTUAL:

Malignant (TP) 151 (FN) 19 170

Benign (FP) 2 (TN) 283 285

All 153 302 455

Classification Report:

precision recall f1-score support

Malignant 0.99 0.89 0.93 170

Benign 0.94 0.99 0.96 285

avg / total 0.96 0.95 0.95 455

Overall Metrics:

accuracy: 0.95

roc\_auc: 0.94

precision: 0.99

recall: 0.89

f1: 0.93

lgrthmc loss: 1.59

mse: 0.05

variance: 0.81

Prediction on training set took 0.009284 seconds.

- - - TEST REPORT - - -

Confusion Matrix:

PREDICTED: Malignant Benign All

ACTUAL:

Malignant (TP) 32 (FN) 10 42

Benign (FP) 3 (TN) 69 72

All 35 79 114

Classification Report:

precision recall f1-score support

Malignant 0.91 0.76 0.83 42

Benign 0.87 0.96 0.91 72

avg / total 0.89 0.89 0.88 114

Overall Metrics:

accuracy: 0.89

roc\_auc: 0.86

precision: 0.91

recall: 0.76

f1: 0.83

lgrthmc loss: 3.94

mse: 0.11

variance: 0.53

Prediction on test set took 0.002248 seconds.