Jaswanth Naidu Gade(11633808)
Neural Network
Assignment 2

1. Classify using K-NN Algorithm with K=5.

    Solution:

    Confusion Matrix(ALL being the positive class):

True Label	Predicted Label	
	Yes	No
Yes	TP = 43	FN = 1
No	FP = 4	TN = 24

$$True\ Positive\ Rate = \frac{TP}{TP+FN} = \frac{43}{43+1} = 0.9772$$

$$False\ Positive\ Rate = \frac{FP}{FP+TN} = \frac{4}{4+24} = 0.1428$$

Confusion Matrix(AML being the positive class):

True Label	Predicted Label	
	Yes	No
Yes	TP = 24	FN = 4
No	FP = 1	TN = 43

$$True\ Positive\ Rate = \frac{TP}{TP+FN} = \frac{24}{24+4} = 0.8571$$

$$False\ Positive\ Rate = \frac{FP}{FP+TN} = \frac{1}{1+43} = 0.0227$$

WEKA(Output for IBK):

Classifier output

```
=== Run information ===

Scheme:      weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""
Relation:     leukemia gene expression data names changed
Instances:   72
Attributes:  151
             [list of attributes omitted]
Test mode:   5-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          67                93.0556 %
Incorrectly Classified Instances         5                 6.9444 %
Kappa statistic                        0.851
Mean absolute error                    0.0839
Root mean squared error                0.2596
Relative absolute error               17.6068 %
Root relative squared error           53.2044 %
Total Number of Instances               72

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
               0.977    0.143    0.915      0.977   0.945      0.854  0.937     0.934     ALL
               0.857    0.023    0.960      0.857   0.906      0.854  0.937     0.888     AML
Weighted Avg.  0.931    0.096    0.932      0.931   0.930      0.854  0.937     0.916

=== Confusion Matrix ===

  a  b   <-- classified as
 43  1 |  a = ALL
  4 24 |  b = AML
```
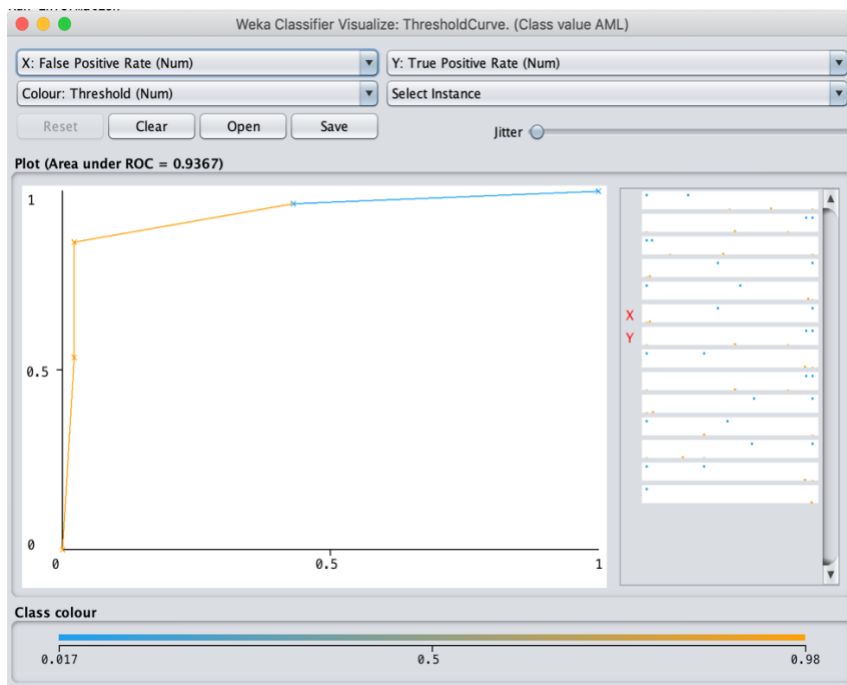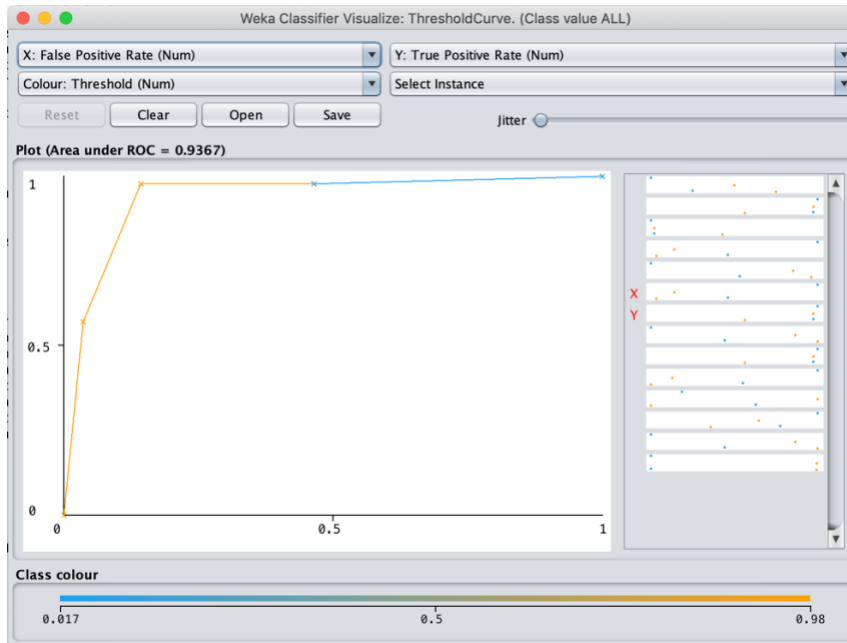
∴ *Calculated TPR and FPR values match with the values generated by WEKA.*

ROC Curves:

2. Classify using ZeroR Algorithm with K folds (K=5).

Solution:

Confusion Matrix(ALL being the positive class):

True Label	Predicted Label	
	Yes	No
Yes	TP = 44	FN = 0
No	FP = 28	TN = 0

$$True\ Positive\ Rate = \frac{TP}{TP + FN} = \frac{44}{44 + 0} = 1.00$$

$$False\ Positive\ Rate = \frac{FP}{FP + TN} = \frac{28}{28 + 0} = 1.00$$

Confusion Matrix(AML being the positive class):

True Label	Predicted Label	
	Yes	No
Yes	TP = 0	FN = 44
No	FP = 0	TN = 28

$$True\ Positive\ Rate = \frac{TP}{TP + FN} = \frac{0}{0 + 44} = 0.00$$

$$False\ Positive\ Rate = \frac{FP}{FP + TN} = \frac{0}{0 + 28} = 0.00$$

WEKA(Output for ZeroR):

Classifier output

```
=== Run information ===

Scheme:        weka.classifiers.rules.ZeroR
Relation:      leukemia gene expression data names changed
Instances:     72
Attributes:    151
               [list of attributes omitted]
Test mode:     5-fold cross-validation

=== Classifier model (full training set) ===

ZeroR predicts class value: ALL

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          44                61.1111 %
Incorrectly Classified Instances        28                38.8889 %
Kappa statistic                          0
Mean absolute error                      0.4765
Root mean squared error                  0.4879
Relative absolute error                100        %
Root relative squared error            100        %
Total Number of Instances               72

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
                1.000    1.000    0.611      1.000   0.759      ?      0.469     0.596     ALL
                0.000    0.000    ?          0.000   ?          ?      0.469     0.373     AML
Weighted Avg.   0.611    0.611    ?          0.611   ?          ?      0.469     0.509

=== Confusion Matrix ===

  a  b   <-- classified as
 44  0 |  a = ALL
 28  0 |  b = AML
```
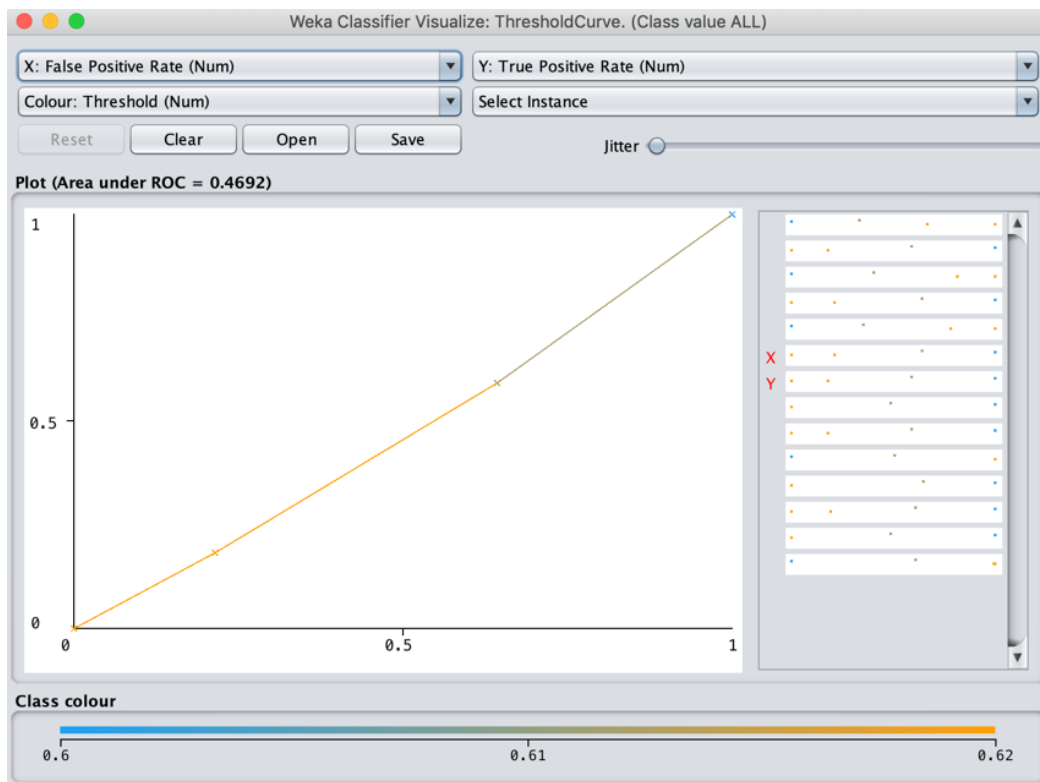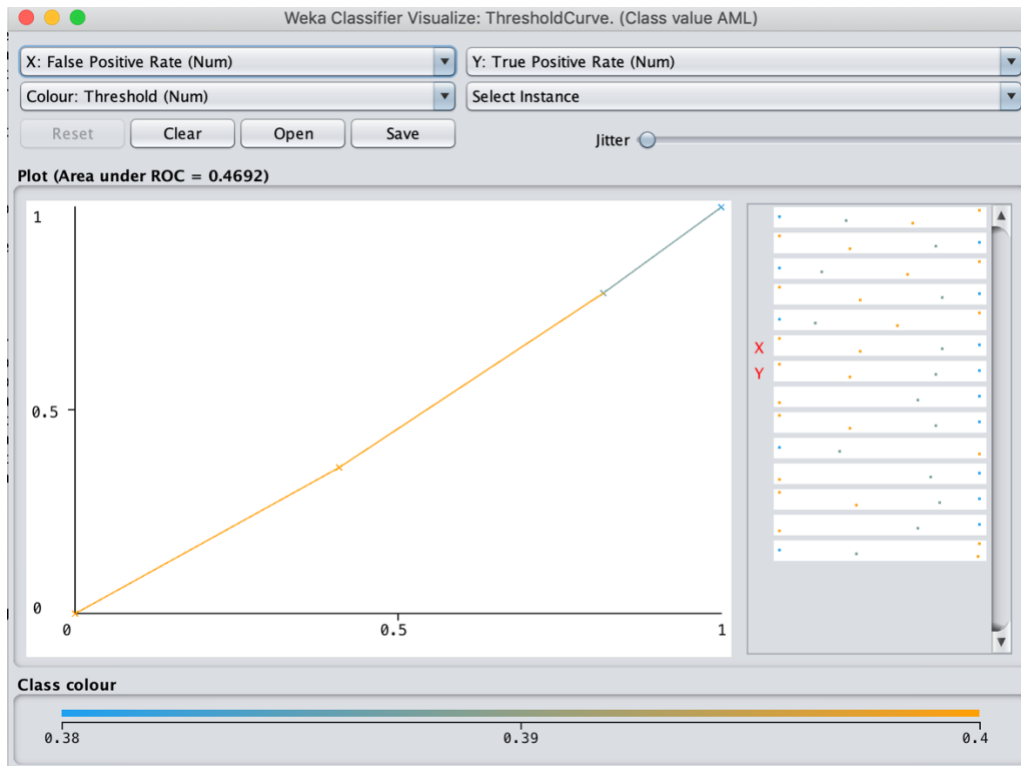
∴ *Calculated TPR and FPR values match with the values generated by WEKA.*

ROC Curves(ZeroR):

Jaswanth Naidu Gade(11633808)
Neural Network
Assignment 2

Result: Instance based classifier (IBK) or KNN outperforms ZeroR method on this dataset because the examples are almost equally spread with both the class labels. ROC curves are determined by the amount of area that lies under the curve in the graph. This area determines the classification capability of the model. IBK has a value of 93% and ZeroR has a value of 46% this clearly specifies that IBK is a better option than ZeroR on this dataset.