**Bayes**

Capture the Confusion Matrix assuming ALL is the positive class

Construct the Confusion Matrix assuming AML is the positive class

Calculate the class-dependent TP and FP rates. Show your work and compare to Weka.

Compare with results from HW 2

Correctly Classified Instances 68 94.4444 %

TP Rate FP Rate Class

1.000 0.143 ALL

0.857 0.000 AML

|  |  |
| --- | --- |
|  |  |

=== Confusion Matrix ===

ALL as the positive class (Weka):

a b <-- classified as

44 0 | a = ALL

4 24 | b = AML

AML as the positive class (Derived):

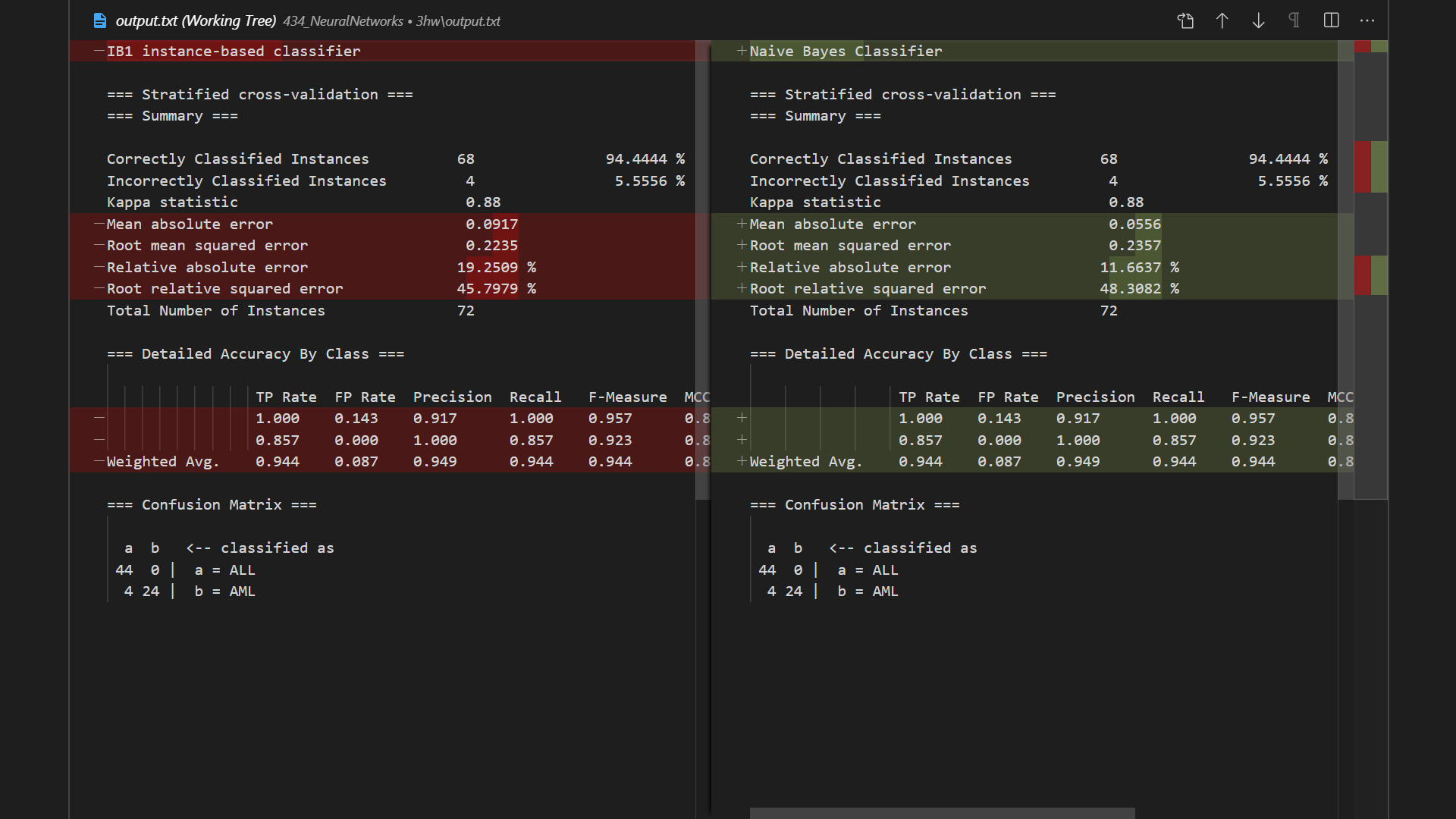
b a <-- classified as

24 4 | b = AML

0 44 | a = ALL

My calculated values match the Weka results.

The Bayes confusion matrix is equal to the KNN confusion matrix from HW2.



|  |  |
| --- | --- |
| **KNN** | **Bayes** |
| ROC Area PRC Area Class  0.976 0.978 ALL  0.976 0.967 AML  Weighted Avg. 0.976 0.973 | ROC Area PRC Area Class  0.942 0.933 ALL  0.978 0.976 AML  Weighted Avg. 0.956 0.949 |

**Bayes**

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes

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) ===

Naive Bayes Classifier

Class

Attribute ALL AML (0.61) (0.39)

…

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 68 94.4444 %

Incorrectly Classified Instances 4 5.5556 %

Kappa statistic 0.88

Mean absolute error 0.0556

Root mean squared error 0.2357

Relative absolute error 11.6637 %

Root relative squared error 48.3082 %

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 0.143 0.917 1.000 0.957 0.886 0.942 0.933 ALL

0.857 0.000 1.000 0.857 0.923 0.886 0.978 0.976 AML

Weighted Avg. 0.944 0.087 0.949 0.944 0.944 0.886 0.956 0.949

=== Confusion Matrix ===

a b <-- classified as

44 0 | a = ALL

4 24 | b = AML

**KNN**

=== Run information ===

Scheme: weka.classifiers.lazy.IBk -K 5 -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 5 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 68 94.4444 %

Incorrectly Classified Instances 4 5.5556 %

Kappa statistic 0.88

Mean absolute error 0.0917

Root mean squared error 0.2235

Relative absolute error 19.2509 %

Root relative squared error 45.7979 %

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 0.143 0.917 1.000 0.957 0.886 0.976 0.978 ALL

0.857 0.000 1.000 0.857 0.923 0.886 0.976 0.967 AML

Weighted Avg. 0.944 0.087 0.949 0.944 0.944 0.886 0.976 0.973

=== Confusion Matrix ===

a b <-- classified as

44 0 | a = ALL

4 24 | b = AML