**CPT\_S 534 HW1**

**Yang Zhang**

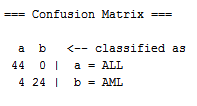
**11529139**

**Question 1**.  What is the mean value of expression of the gene labeled “CD33 CD33 antigen (differentiation antigen)”?

The mean value = **1884.514**

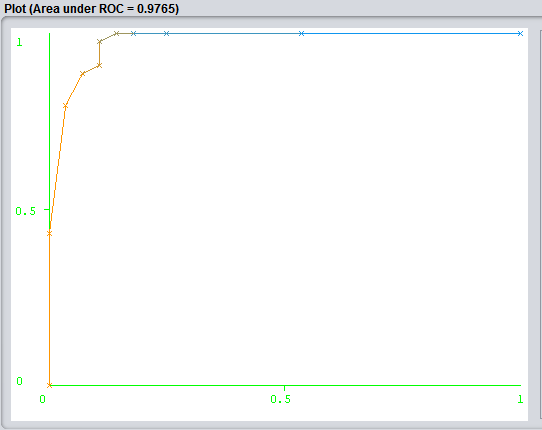
**KNN Classifier**

Confusion Matrix:

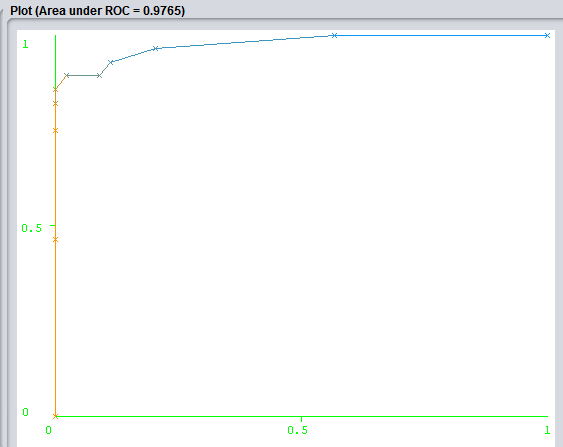


Roc area: RALL **=** 0.9765**,** RAML **=** 0.9765

ALL class ROC curve:



AML class ROC curve:



**Question 2a**.  What is the % of correctly classified instances?

**94.4444 %**

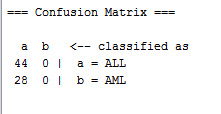
**Question 2b**.  What are the TP and FP rates for ALL and AML?

**TPALL = 1.000, FPALL = 0.143**

**TPAML = 0.857, FPAML = 0.000**

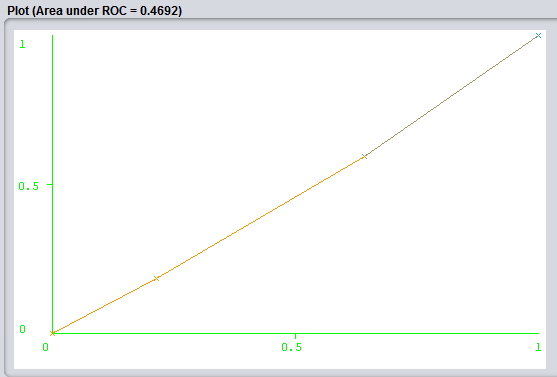
**ZeroR Classifier**

Confusion Matrix:

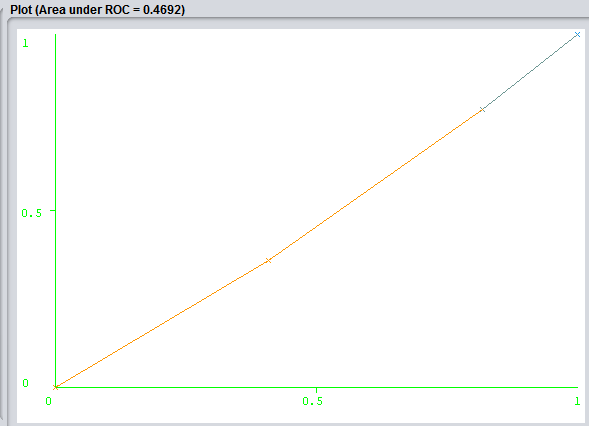


Roc area: RALL **=** 0.4692**,** RAML **=** 0.4692

ALL class ROC curve:



AML class ROC curve:



**Question 3a**: What is the % of correctly classified instances?

**61.1111%**

**Question 3b**: What are the TP and FP rates for ALL and AML?

**TPALL = 1.000, FPALL = 1.000**

**TPAML = 0.000, FPAML = 0.000**

**Question 3c**: Explain your 3b results from the definitions of TP and FP rates.

For ALL, TP = 1.000 means that all cases that predicted yes are actually yes.

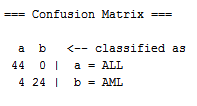
FP = 1.000 means that all cases that predicted yes are actually no.

For AML, TP = 0.000 means that no case that predicted yes are actually yes.

FP = 0.000 means that no case that predicted yes are actually no.

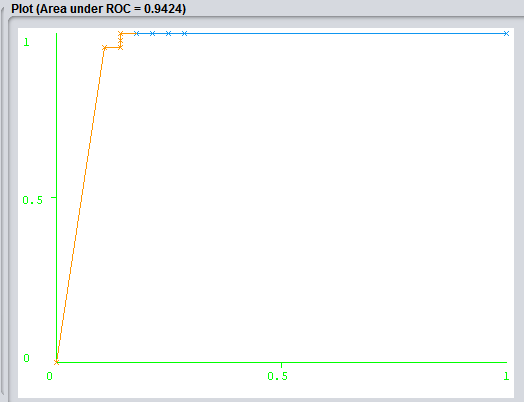
**NaiveBayes Classifier**

Confusion Matrix:

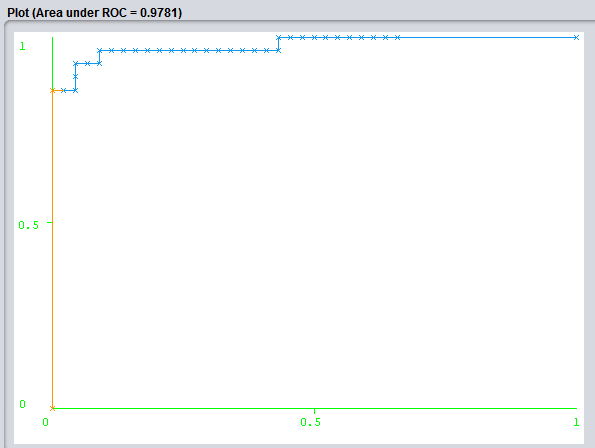


Roc area: RALL **=** 0.9424**,** RAML **=** 0.9781

ALL class ROC curve:



AML class ROC curve:



**Question 4a**: What is the % of correctly classified instances?

94.4444 %

**Question 4b**: What are the TP and FP rates for ALL and AML?

**TPALL = 1.000, FPALL = 0.143**

**TPAML = 0.857, FPAML = 0.000**

1. The objective of this datamining exercise is to familiar with data analysis tool Weka and apply different classifiers to the same data set to observe the difference and understand the way to judge a good class based on the result.
2. My assessment of class’s success based on its ROC area and its TP and FP rates.
3. AML class is better, since in general, AML class has the bigger ROC area.