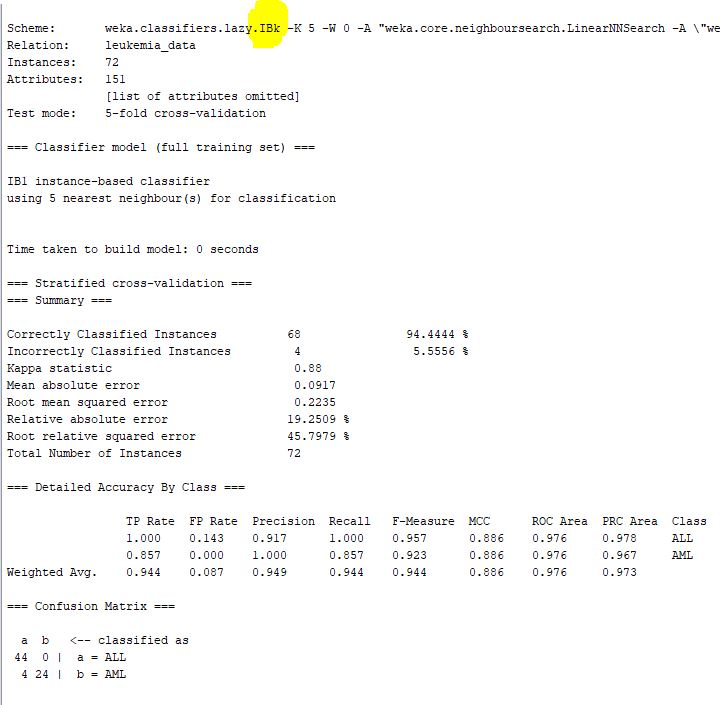
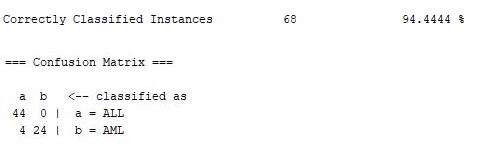
# Type of Leukemia Prediction

## Part 1 - Using Weka



### 1) Record % correctly classified and confusion matrix (ALL positive class).



**True Positive Rate** = TP / (TP + FN) = 44 / (44 + 0) = 1.0

100% sensitive

**False Positive Rate** = FP / (FP + TN) = 4 / (4 + 24) = 0.143

low error

### What are ALL and AML stand for?

Types of Leukemia:

. **ALL = Acute lymphocytic leukemia**

. **AML = Acute myelocytic leukemia**

### 2) Derive the confusion matrix when AML is the positive class.

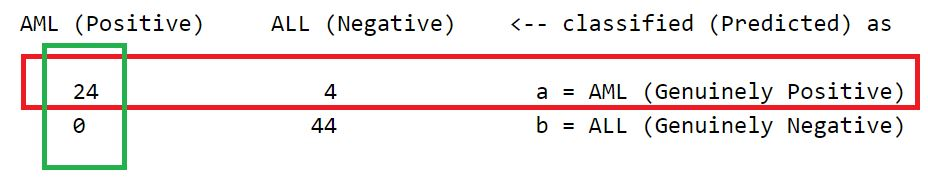
AML (Positive) ALL (Negative) <-- classified (Predicted) as  
  
 24 4 a = AML (Genuinely Positive)  
 0 44 b = ALL (Genuinely Negative)

### 3) Calculate the class-dependent TP and FP rates when AML is the positive class

### TPR (Sensitivity): (more is better)

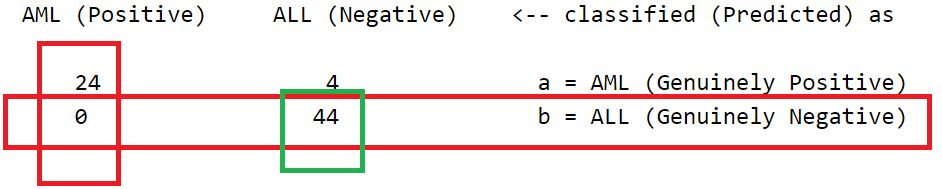
True Positive Rate = TP / (TP + FN) = 24 / (24 + 4) = 0.857

85.7% Sensitive



### FPR (1 - Specificity): (less is better) (less = more specific)

False Positive Rate = FP / (FP + TN) = 0 / (0 + 44) = 0



### ROC curves

ROC = Receiver Operator Characteristic

. Helps determine the cutoff point which optimizes sensitivity and specificity for given tests.

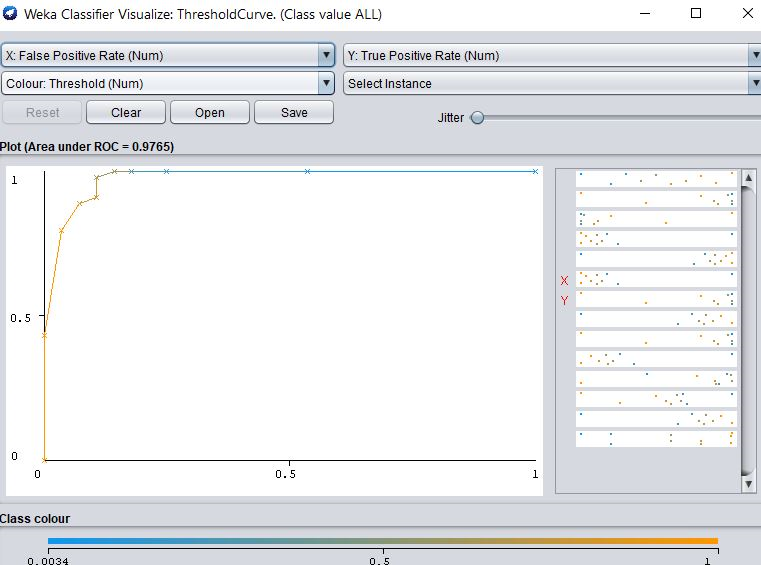
. Can be used to assess the overall diagnostic accuracy of a test

. Y axis = TPR = sensitivity (more is better)

. X axis = FPR = (1 - specificity) (less is better)

. Overall Diagnostic Accuracy = AUR = Area Under ROC curve

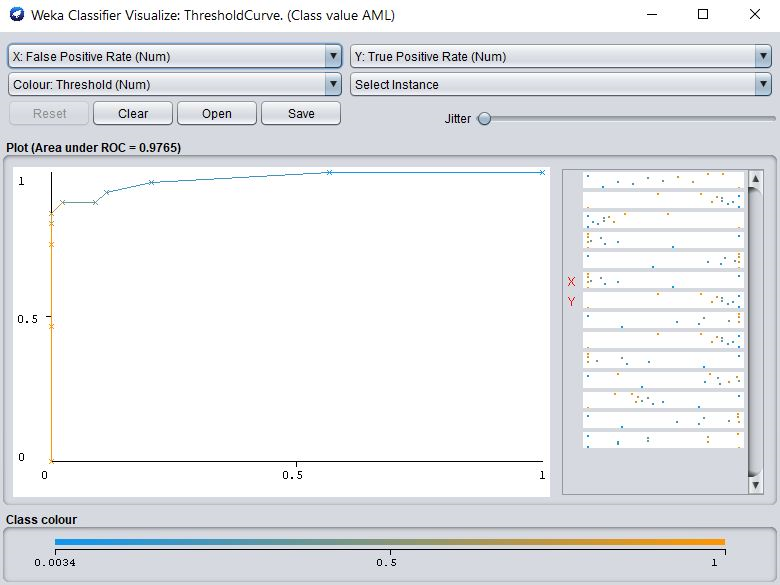
### 4) Capture the ROC curves for ALL positive



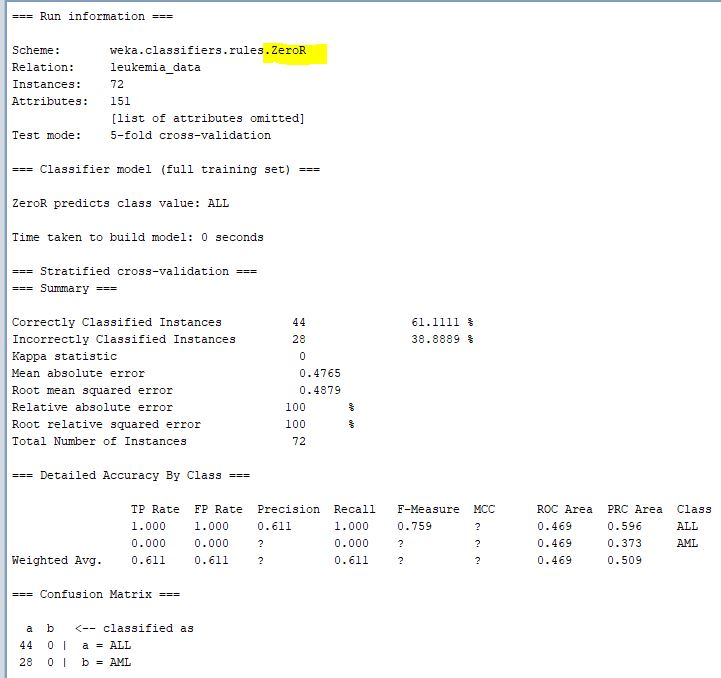
### 

### 5) Capture the ROC curves for AML positive

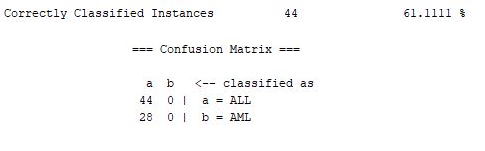
ROC = Receiver Operator Characteristic



### 6) Using the ZeroR baseline classifier



### 7) Record % correctly classified and confusion matrix (ALL positive class).



**True Positive Rate** = TP / (TP + FN) = 44 / (44 + 0) = 1.0

100% sensitive

**False Positive Rate** = FP / (FP + TN) = 28 / (28 + 0) = 1.0

100% erroneous

### 8) Derive the confusion matrix when AML is the positive class.

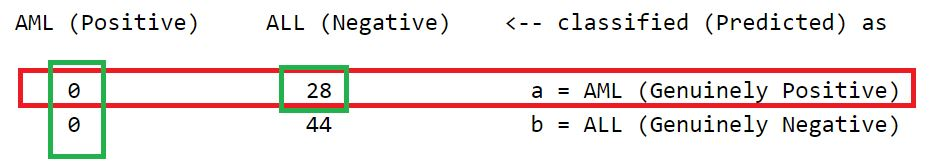
AML (Positive) ALL (Negative) <-- classified (Predicted) as  
  
 0 28 a = AML (Genuinely Positive)  
 0 44 b = ALL (Genuinely Negative)

### 9) Calculate the class-dependent TP and FP rates when AML is the positive class

### TPR (Sensitivity): (more is better)

True Positive Rate = TP / (TP + FN) = 0 / (0 + 28) = 0

0% Sensitive

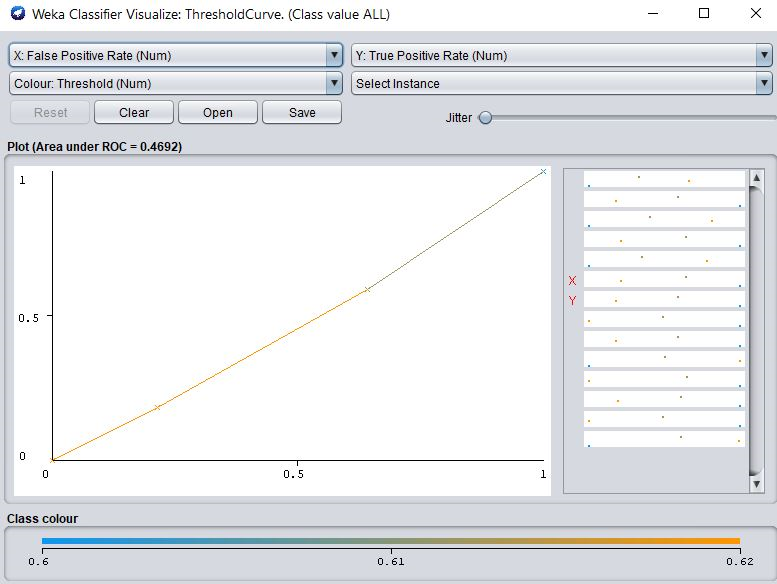


### FPR (1 - Specificity): (less is better) (less = more specific)

False Positive Rate = FP / (FP + TN) = 0 / (0 + 44) = 0

### 10) Capture the ROC curves for ALL positive

ROC = Receiver Operator Characteristic



### 11) Capture the ROC curves for AML positive

ROC = Receiver Operator Characteristic

