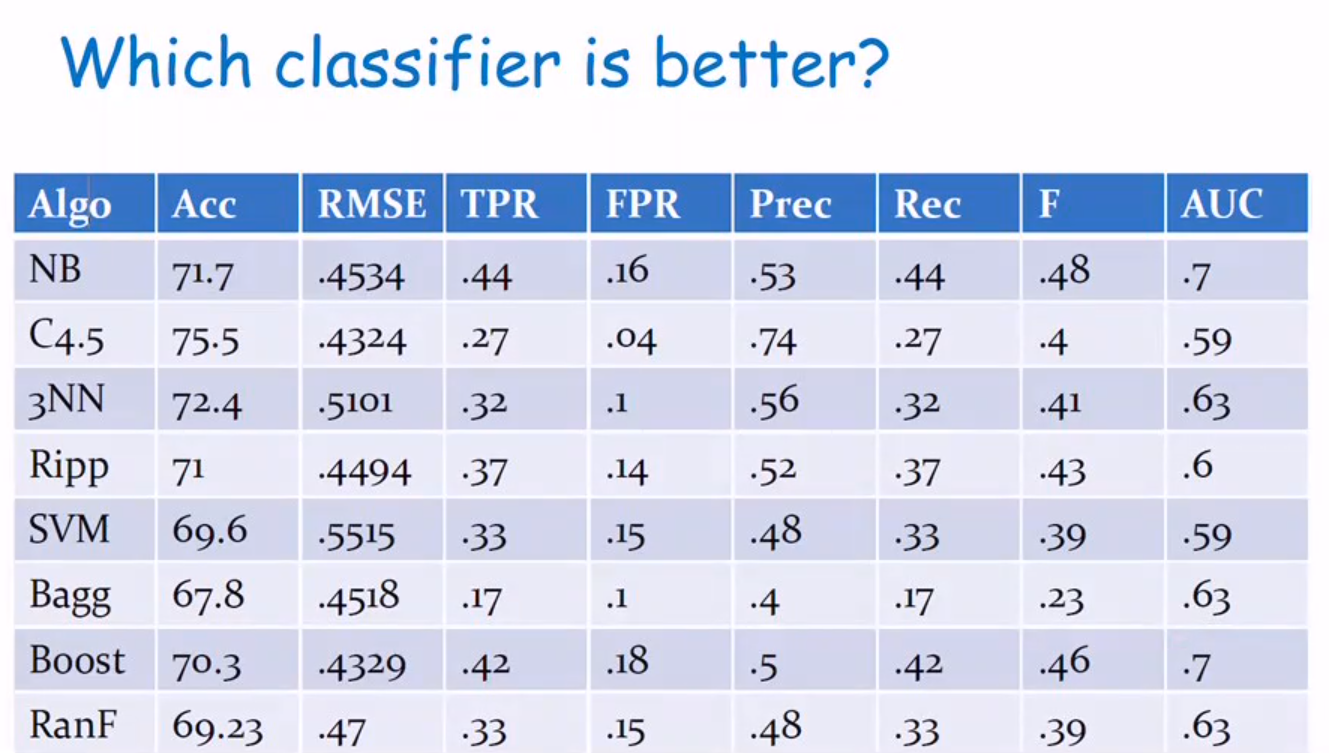
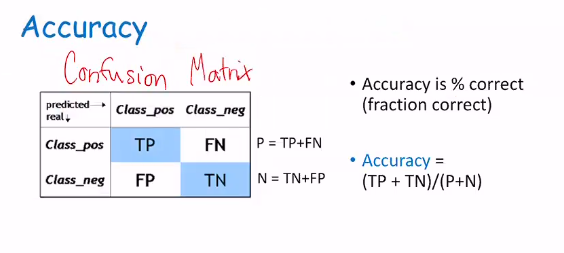
**Machine Learning : Evaluation**

**Performance Measures**

**What classifier is better?**



**Accuracy**

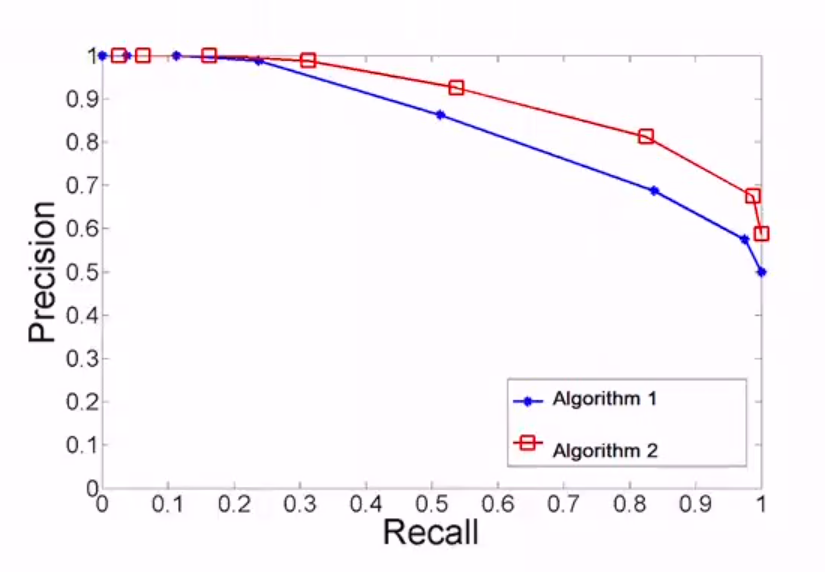
* **TP (True Positive) means the classifier got it right and the ground truth value is positive**
* **TN (True Negative) means the classifier got it right and the ground truth value is negative**
* **FN (False Negative) means the classifier got it wrong when the ground truth value is positive**
* **FP (False Positive) means the classifier got it wrong when the ground truth value is negative**
* **Ground truth is the real value**

**Issues with Accuracy**

* **Label of positive and negative might make a difference**

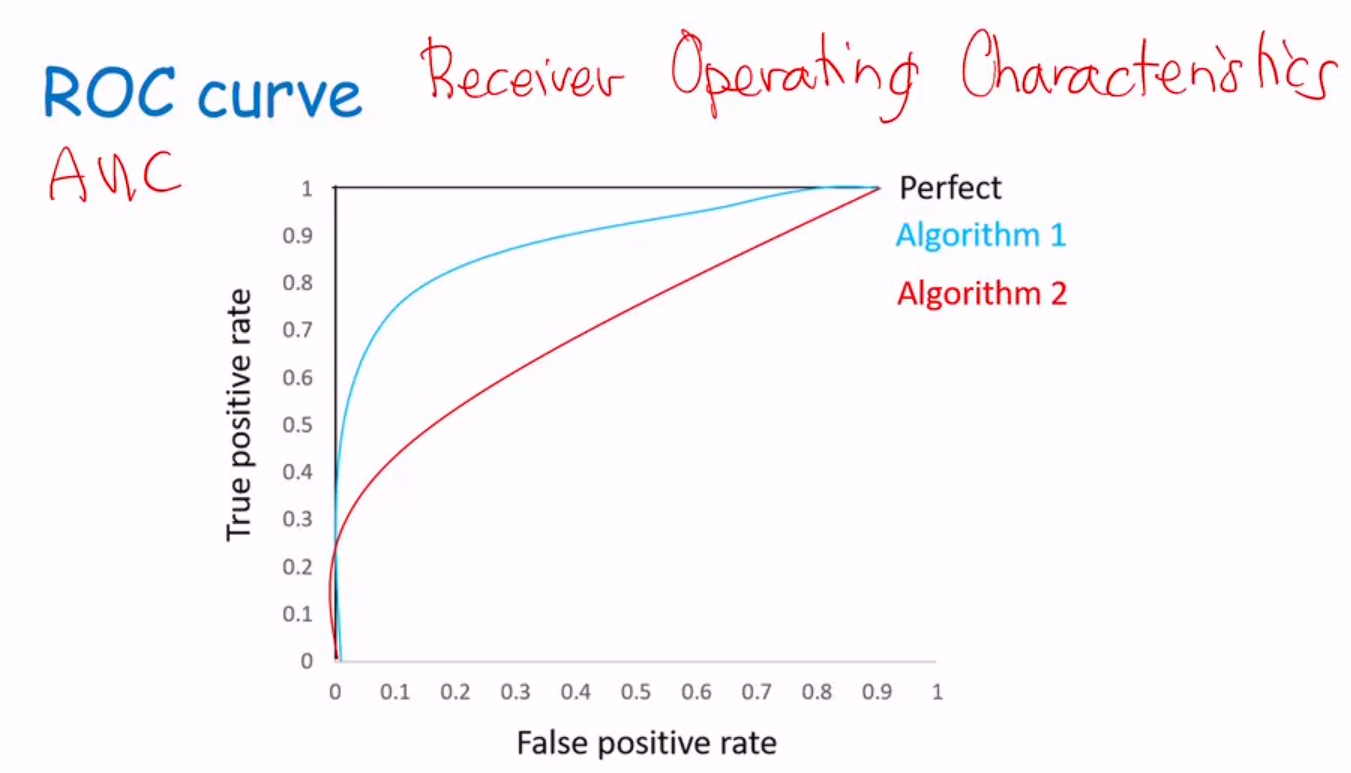
**Discrimination vs detection (spotting)**

* **Precision and Recall**
  + **Precision : of all the target classes activities how, many were actually that activity**
    - **Precision = TP/(TP+FP)**
  + **Recall : of all the activities that exists how many did you find**
    - **Recall = TP/(TP+FN)=TP/P**
* **Spot instances of a pattern, and only want the pattern we are interested in**



* **You want both recall and precision to be very close to 1**
* **Algorithm with a higher area under the curve performs better, so algorithm 2 performs better**

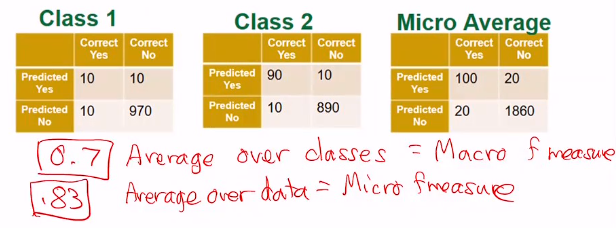
**ROC (Receiver operating characteristics) Curve**



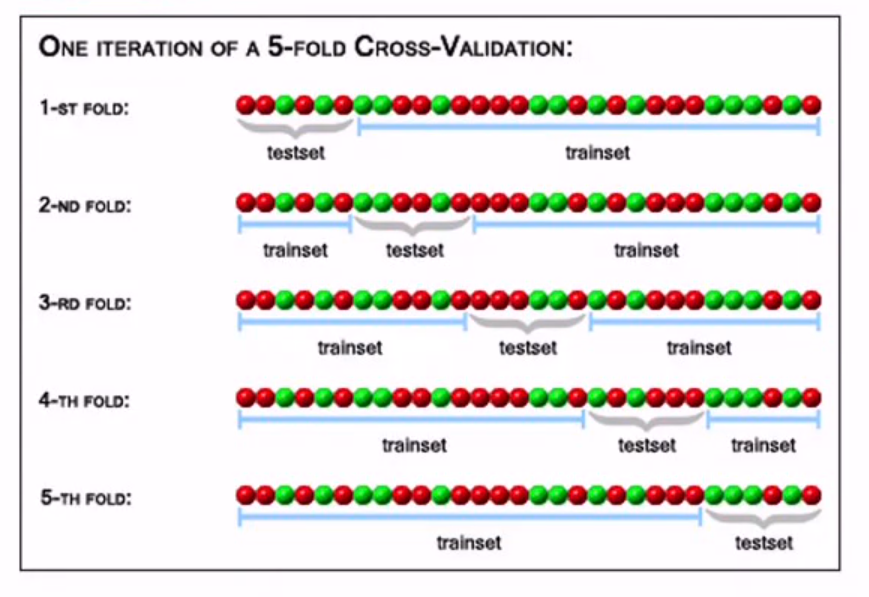
* **AUC – Area under the ROC curve**
* **True Positive Rate (TPR) = TP/(TP + FN)**
* **False Positive Rate (FPR) = TN/(TN + FP)**

**F-Measure**

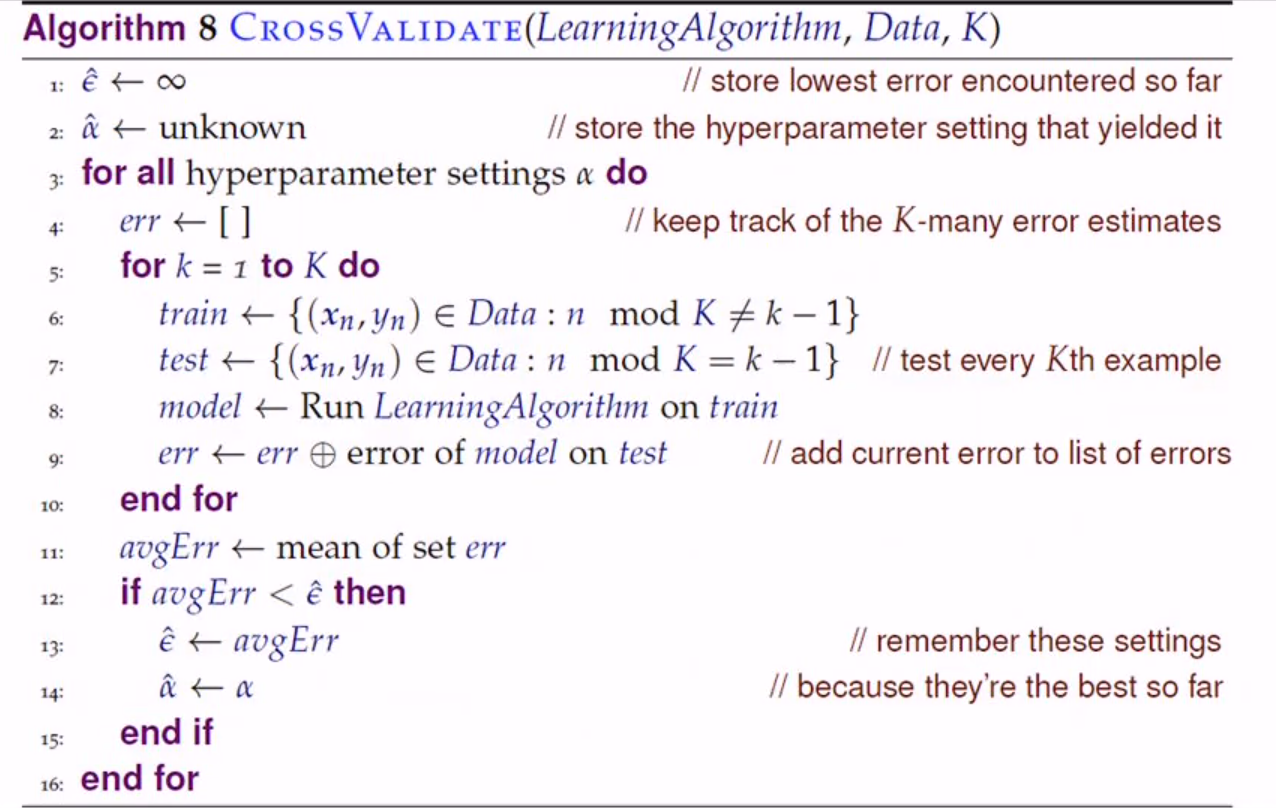
* **F = (2 \* P \* R) / (P + R) where P is precision and R is recall**
* **Combines Precision and Recall**
* **Imbalance between the precision and recall will have a lower f-measure vs a balanced precision and recall of the same sum.**
  + **F(0.6,0.2)= 0.3 vs F(0.4,0.4)=0.4**

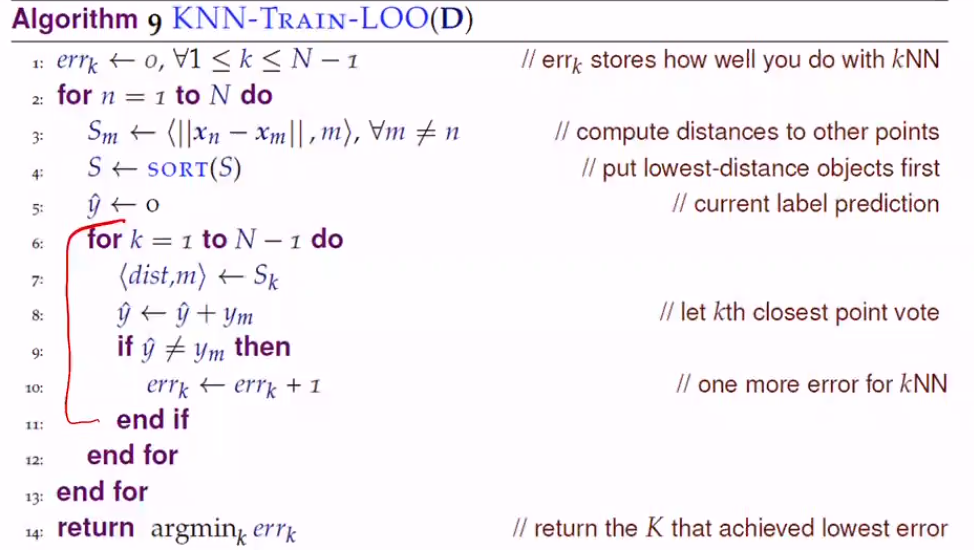


**Cross-Validation**



* **Let every part of your data be test data at some point, by partitioning your data into multiple partitions (folds) and rerun your training and testing as many times as you have folds. Each time through a different partition is your test.**
* **Average over all the folds**

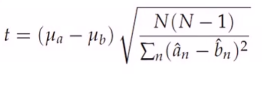




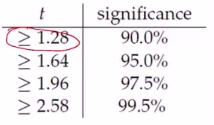
**Statistical significance is high when the p-value is low**

**T-test**

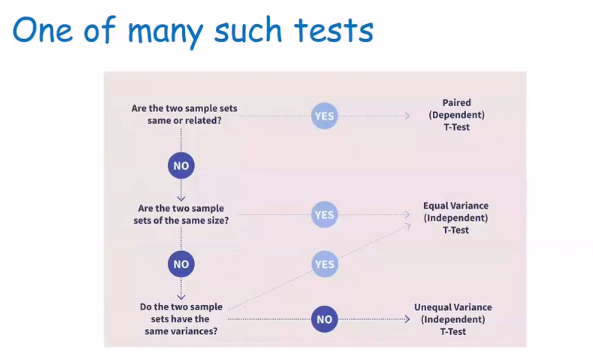
* Compute p-value
* Probability that observed difference was luck
* Parametric – parameters, assumptions
  + Assumptions such as data follows gaussian distribution
* Means of the two groups are equal
* Calculate sample mean (population mean)
* Degrees of freedom = n-1
  + =#of data points
* Error of algorithm
* Error of algorithm B is
* Center data points around means
  + Each is now
  + Each is now



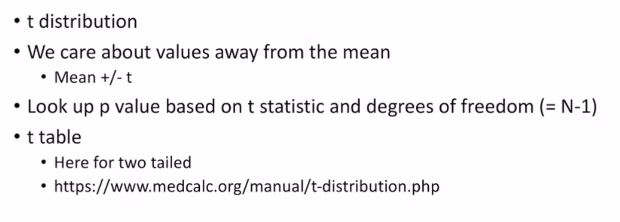
* t-value not restricted
* p value [0…1]
  + represents value of t s.t. probability (T>t)>p

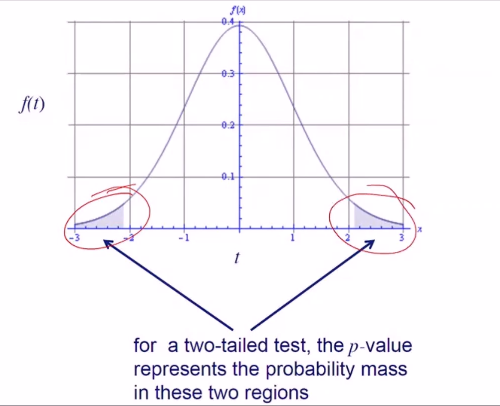


* Greater than t value the greater the significance



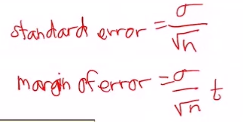
**Calculating p-value form t statistic**

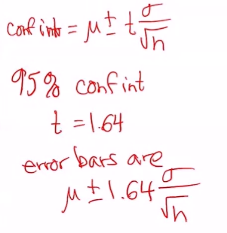


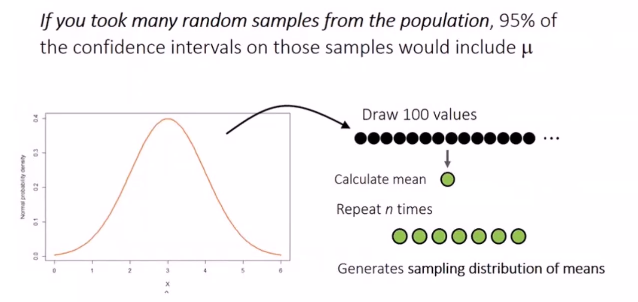


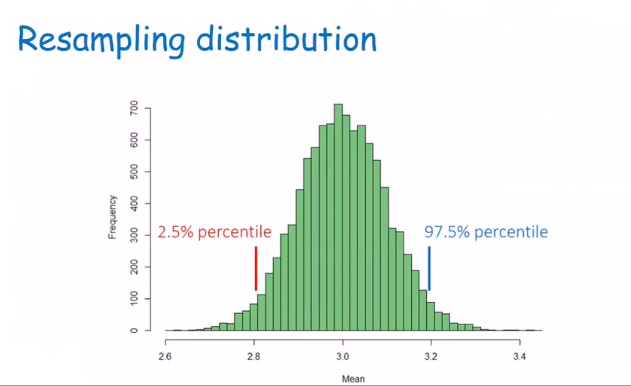
**Confidence Intervals**

* **The median for the population lies between**
* **As population variation becomes small the confidence interval becomes small**
* **Population variation increases confidence interval increases**
* **Sample size decreases confidence interval increases**
* **Larger sample confidence interval decreases**

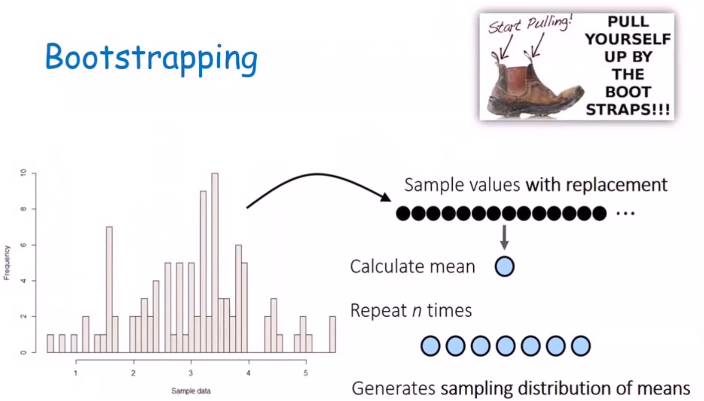


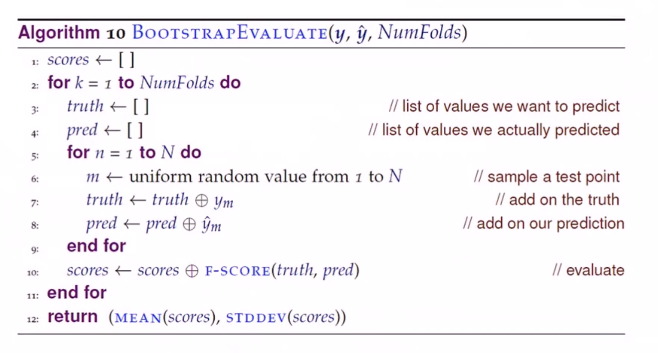






**Bootstrapping**





**Learning Curve**



* **Good to know that your model is learning something**
* **If it is learning it should have this shape**