INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD
OUTREACH DIVISION

talent sprint

Special Lecture 4

Accuracy, Precision And Recall

Performance Evaluation Metrics

Accuracy Metrics

Revisiting Binary case...

$$Accuracy = \frac{(100+50)}{165} = 0.91$$

$$Misclassification = \frac{(10+5)}{165} = 0.09$$

$$TruePositiveRate(TP) = \frac{(100)}{105} = 0.95$$

$$FalsePositiveRate(FP) = \frac{(10)}{60} = 0.17$$

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Revisiting Binary case...

$$TrueNegativeRate(TN) = \frac{(50)}{60} = 0.833$$

$$FalseNegativeRate(FN) = \frac{(5)}{105} = 0.048$$

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Set notation. Key accuracy measures and terminologies

- Classification Error = $\frac{errors}{total}$

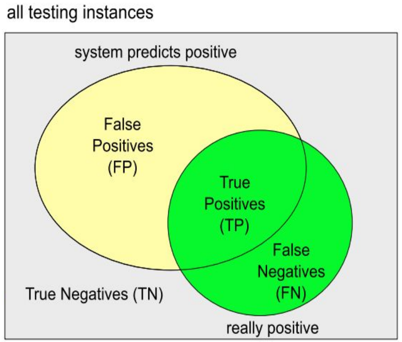
$$= \frac{FP + FN}{TP + TN + FP + FN}$$

- Accuracy = 1 - Error = $\frac{correct}{Total}$

$$= \frac{TP + TN}{TP + TN + FP + FN}$$

Predict positive?

Really positive?	Yes	No
Yes	TP	FN
No	FP	TN



Set notation. Key accuracy measures and terminologies

- False Alarm = False Positive Rate

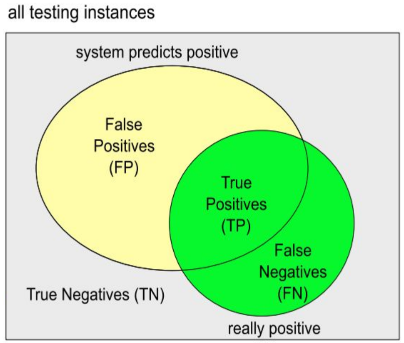
$$= \frac{FP}{TN + FP}$$

- Miss = False Negative Rate

$$= \frac{FN}{TP + FN}$$

Predict positive? (columns) / Really positive? (rows)

	Yes	No
Yes	TP	FN
No	FP	TN



Set notation. Key accuracy measures and terminologies

- Recall = True Positive Rate

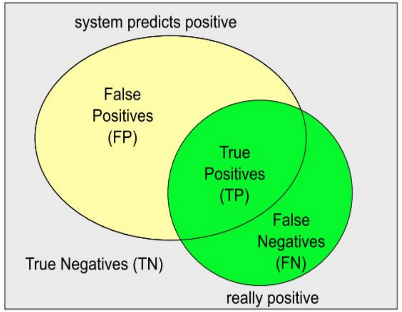
$$= \frac{TP}{TP + FN}$$

- Precision = $\frac{TP}{TP + FP}$

Predict positive?

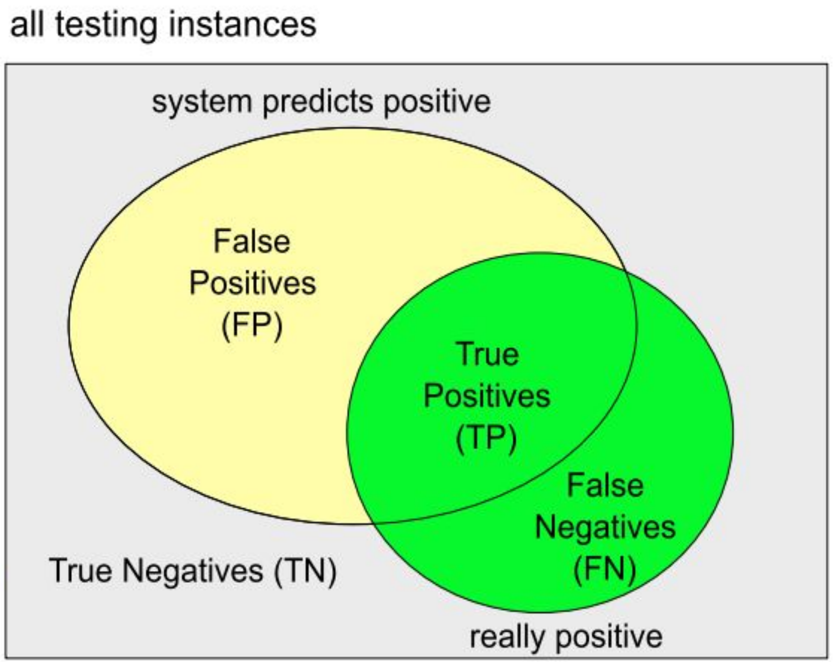
Really positive?	Yes	No
Yes	TP	FN
No	FP	TN

all testing instances



Set notation. Key accuracy measures and terminologies

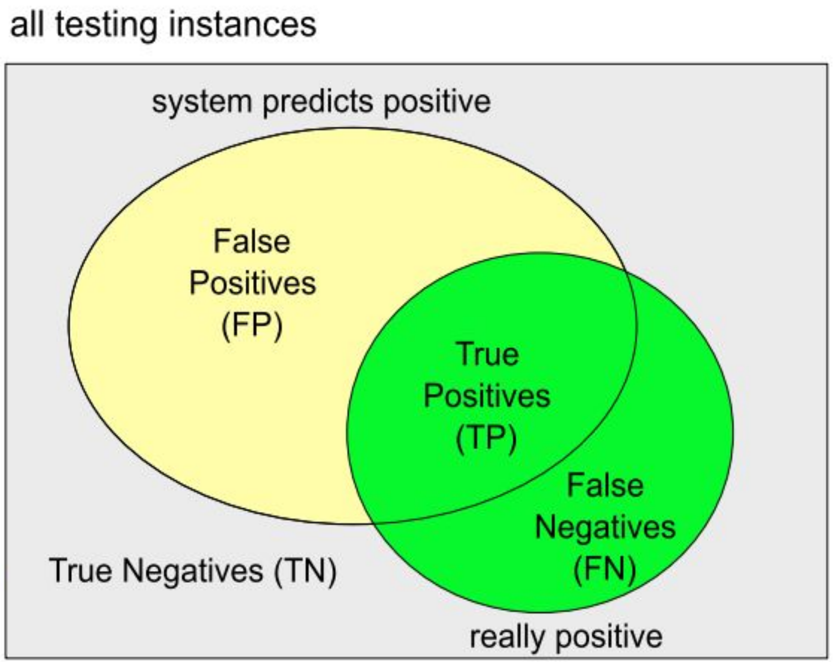
- True Positive Rate also called " Sensitivity"
- "Specificity" = 1- False Alarm



Set notation. Key accuracy measures and terminologies

- “Sensitivity” = Probability of a positive test given a patient has the disease

- “Specificity” = Probability of negative test given a patient is well



Multi-class problems - Confusion matrix

For multi-class problem?

Confusion Matrix

activity recognition from video

actual class

	bend	jack	jump	pjump	run	side	skip	walk	wave1	wave2
bend	100	0	0	0	0	0	0	0	0	0
jack	0	100	0	0	0	0	0	0	0	0
jump	0	0	89	0	0	0	11	0	0	0
pjump	0	0	0	100	0	0	0	0	0	0
run	0	0	0	0	89	0	11	0	0	0
side	0	0	0	0	0	100	0	0	0	0
skip	0	0	0	0	0	0	100	0	0	0
walk	0	0	0	0	0	0	0	100	0	0
wave1	0	0	0	0	0	0	0	0	67	33
wave2	0	0	0	0	0	0	0	0	0	100

predicted class

Courtesy: vision.jhu.edu

Utility and Cost

- Sometimes, there is a cost for each error
  - E.g. Earthquake prediction
    - False positive: Cost of preventive measures
    - False negative: Cost of recovery

- Detection Cost (Event detection) -Can be applied to example above
  - $\text{Cost} = C_{FP} * FP + C_{FN} * FN$

Utility and Cost

- What to do when one classifier has better Precision but worse Recall, while other classifier behaves exactly opposite?
  - F-measure (Information Retrieval)
    - $F_1 = \dfrac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$

Revisiting scenarios where metrics are appropriate

- When you do cancer screening what do you care?
  - High TP
- When you classify between “apple” and “orange”
  - High Accuracy or High TP or High TN
- Automatic Firing on detecting a violation.
  - Very low FP

Precision and Recall

Problem of Retrieval

- You give a query q
- You get a ranked list of documents (say 10)
    - $d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9, d_{10}$
- The document di could be “relevant” (+) or “irrelevant”(-)

Problem of Retrieval

- Let us assume the relevances are:
    - +, +, -, -, +, -, +, -, -, +

Precision and Recall

- Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

$$Precision = \frac{TP}{(TP + FP)}$$

- Recall is the ratio of correctly predicted positive observations to the all observations in actual class

$$Recall = \frac{TP}{(TP + FN)}$$

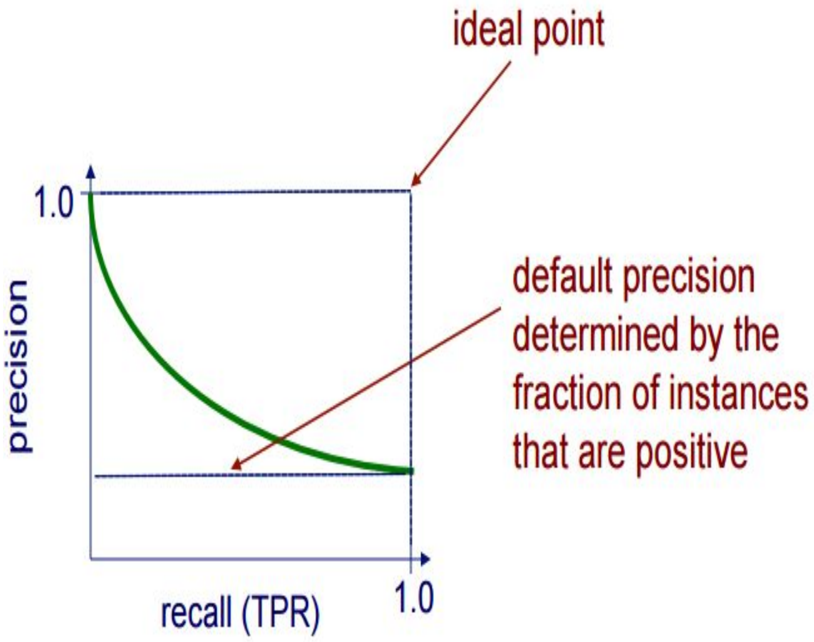
Precision and Recall

Assume there were 10 “True”/”Relevant” documents (often we do not know this)

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}
	+	+	+	+	+	+	+	+	+	+
P@K	1.0	1.0	0.66	0.50	0.60	0.50	0.57	0.50	0.44	0.50
R@K	0.1	0.2	0.2	0.2	0.3	0.3	0.4	0.4	0.4	0.5

Precision/Recall Curves

- A precision/recall curve plots the precision vs. recall (TP-rate) as a threshold on the confidence of an instance being positive is varied



Numerical problem: Precision and recall

- Suppose there are 6000 images of Amitabh Bachchan, ever, on the web. Suppose you fire an image search which is programmed to return 4000 images. Out of this you find 3000 are indeed Amitabh’s images. What the precision and recall in this case?

Solution: Precision and recall

$$Precision = \frac{TP}{(TP+FP)} \quad Recall = \frac{TP}{(TP+FN)}$$

Total images returned = 4000
TP= All the images of Amitabh successfully returned =3000
FP= Images returned that are not Amitabh = 4000-1000
FN=All the images of Amitabh not returned = 6000-3000 = 3000

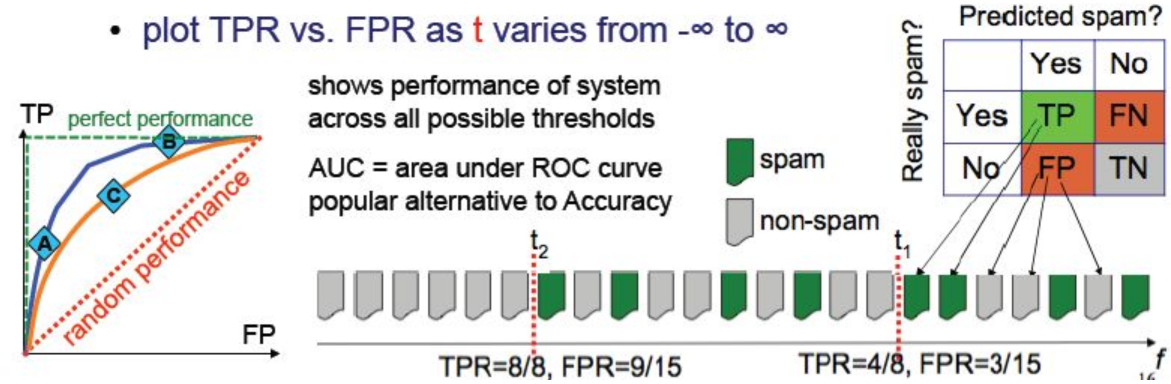
$$Precision = \frac{3000}{3000+1000} = 0.75 \quad Recall = \frac{3000}{3000+3000} = 0.5$$

Classifier Evaluation

Receiver Operating Characteristic

ROC Curves

- Many algorithms compute “confidence” f(x)
  - Threshold to get decision: spam if f(x) > t, non-spam if f(x) <= t
  - Threshold to determines error rates
- Receiver Operating Characteristic (ROC)

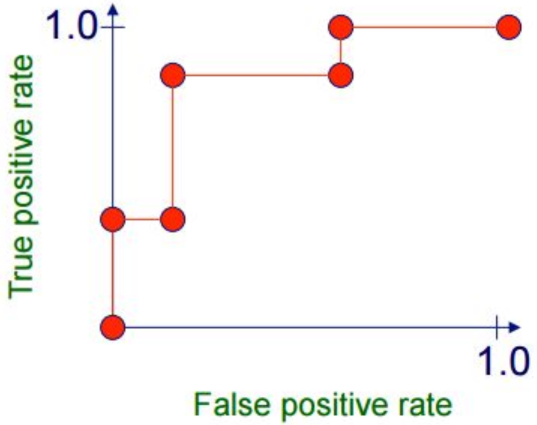


ROC Curve: Algorithm

- Sort test-set predictions according to confidence that each instance is positive
- Step through sorted list from high to low confidence
  - Locate a threshold between instances with opposite classes (keeping instances with the same confidence value on the same side of threshold)
  - Compute TPR, FPR for instances above threshold
  - Output (FPR, TPR) coordinate

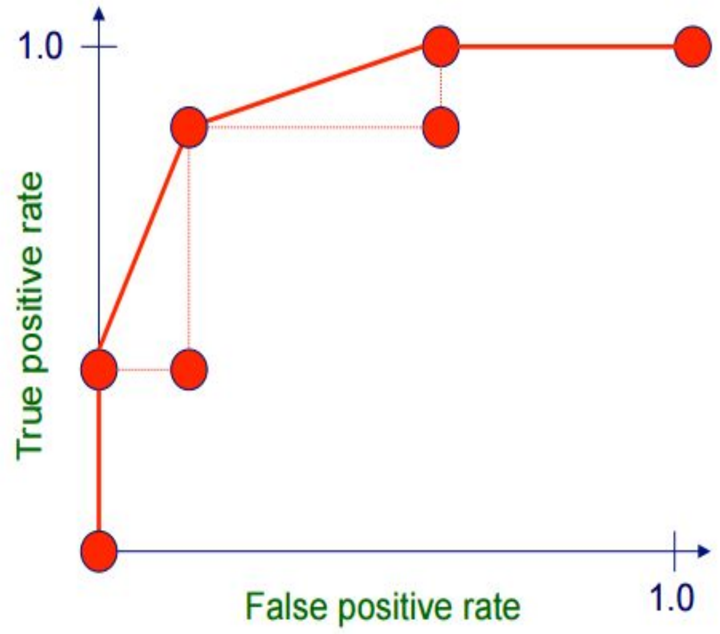
Plotting an ROC Curve

instance	confidence positive		correct class
Ex 9	.99		+
Ex 7	.98	TPR= 2/5, FPR= 0/5	+
Ex 1	.72	TPR= 2/5, FPR= 1/5	-
Ex 2	.70		+
Ex 6	.65	TPR= 4/5, FPR= 1/5	+
Ex 10	.51		-
Ex 3	.39	TPR= 4/5, FPR= 3/5	-
Ex 5	.24	TPR= 5/5, FPR= 3/5	+
Ex 4	.11		-
Ex 8	.01	TPR= 5/5, FPR= 5/5	-

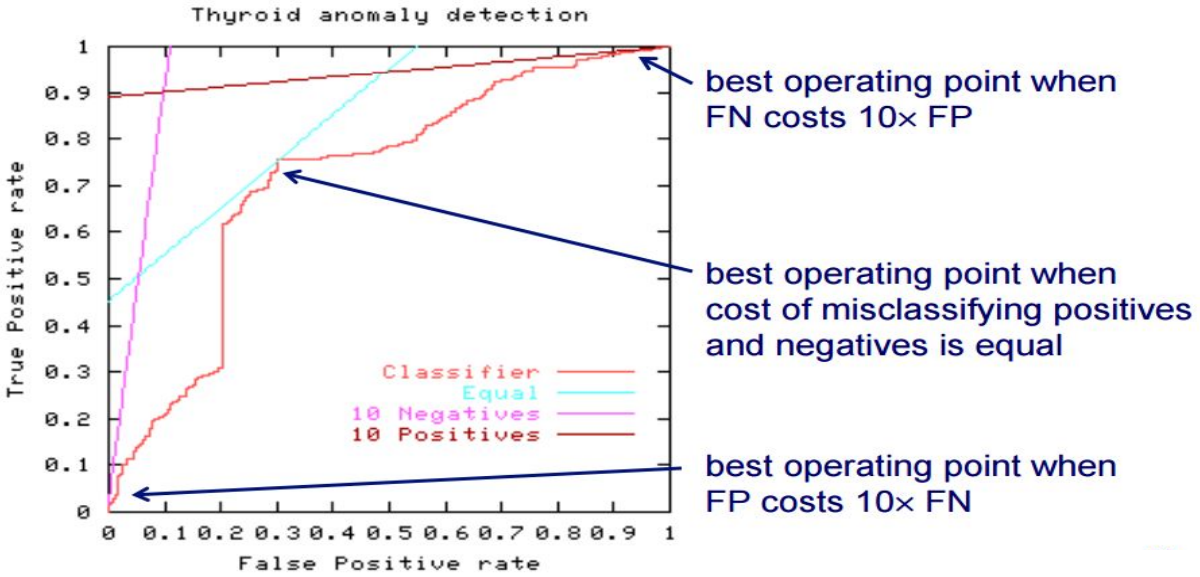


Plotting an ROC Curve

- Can interpolate between points to get convex hull



ROC Curves and Misclassification Costs. Operating Point



Calculating the operating point:

α = cost of a false positive (false alarm)

β = cost of missing a positive (false negative)

p = proportion of positive cases

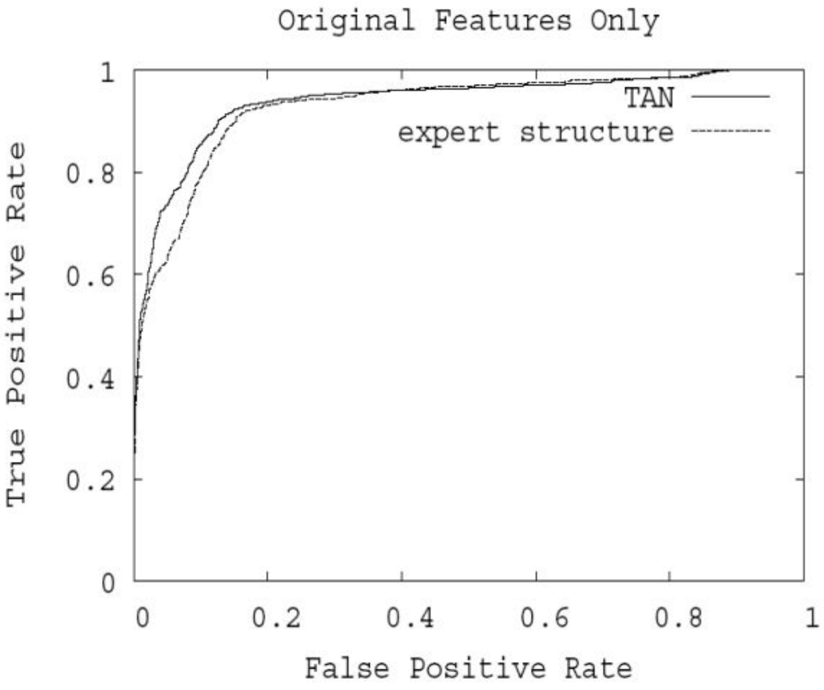
Then the average expected cost of classification at point x,y in the ROC (where x is FP, and y is TP)

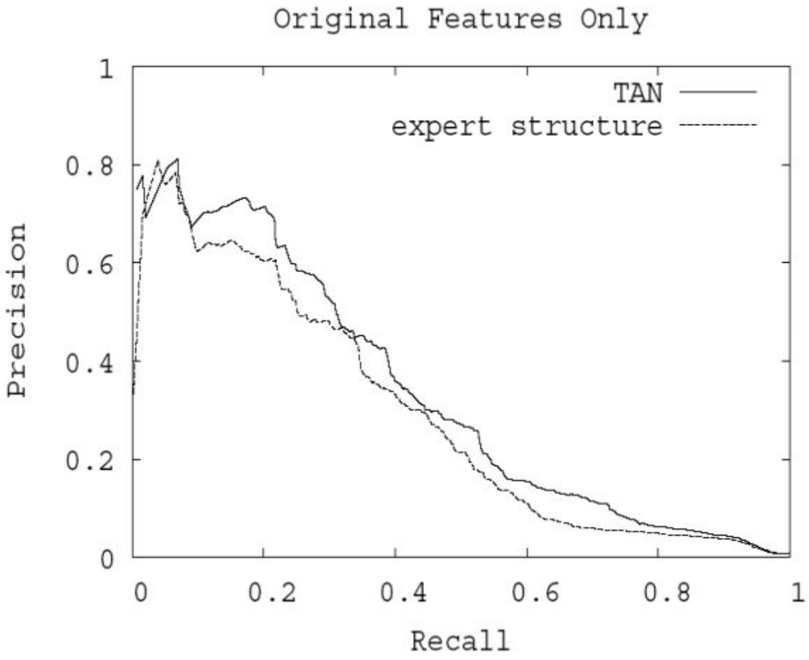
space is

$$c = (1 - p) \times \alpha \times x + p \times \beta \times (1 - y)$$

The blue line in previous slide represents cases α = β.

ROC + PR Curves Example

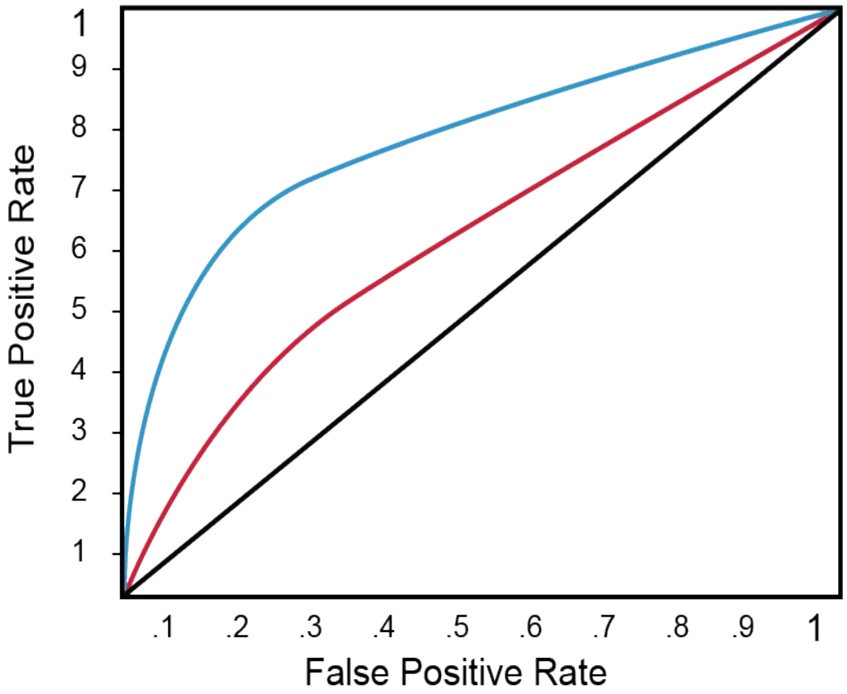




Trade Off...

To compare two screening tests, at ROC(Receiver Operating Characteristics):

The higher the Curve, the better.



Summary

- Many metrics:
  - Accuracy, TP, FP, AUC, Precision, Recall, AP/mAP
- Many problems demand many measures.
  - Choice of right measure is very important.
- Confusion Matrix: Important to analyze and refine solution.
- Curves provide “Trade off” and help to choose operating point.