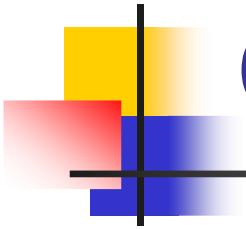
An Introduction to ROC Curve (Receiver Operating Characteristics)

Ming-Chang Lee

Department of Information Management
Yu Da College of Business

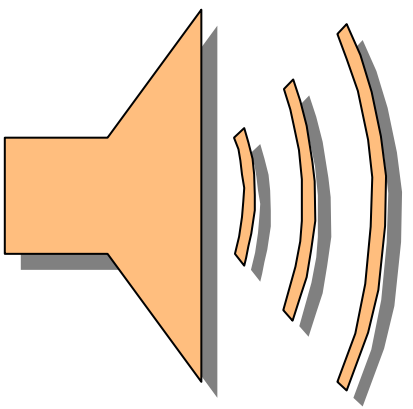


Outline

1. Introduction
2. Create an ROC curve
3. **A**rea **U**nder an ROC **C**urve (AUC)
4. R-package “ROCR” demo
5. References

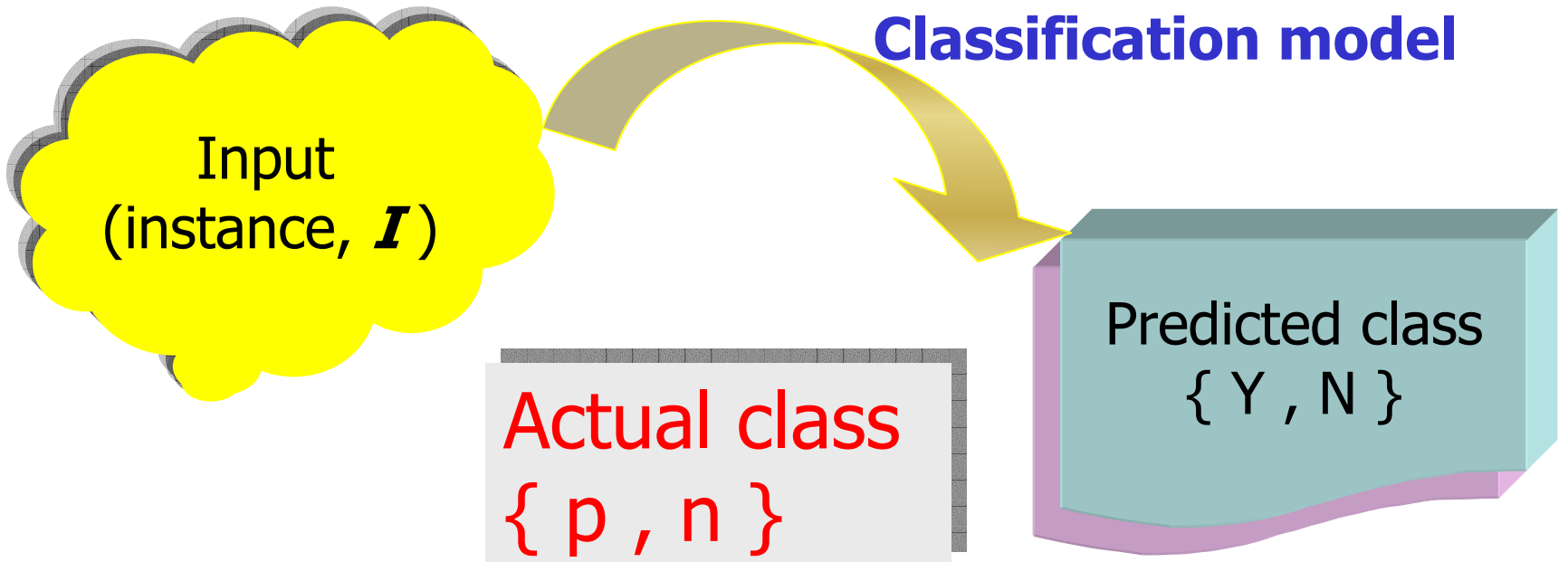
1. Introduction

- History :
  Signal detection theory – hit rates and false alarm rates
- Development:
  - Diagnostic system
  - Medical decision making
  - Machine learning



Classier Performance

- Problem: two classes classification



PS: actual class {p: positive class, n: negative class}

Confusion matrix (Contingency table)

- Given a classifier and an instance:

Classifier / Predicted class	TRUE CLASS	
	p (positive)	n (negative)
Y	True Positives	False Positives
N	False Negatives	True Negatives
Total	P	N

P = True Positives + False Negatives

Performance index

TP	FP
FN	TN

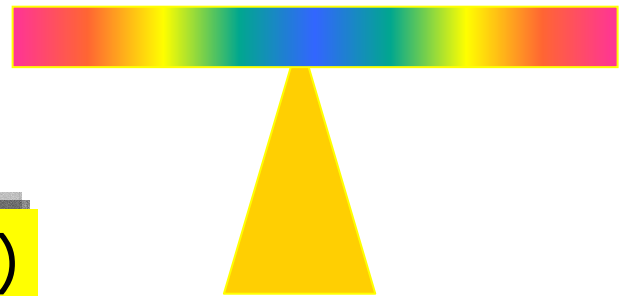
$$TPR = \frac{TP}{P} = Recall,\ FPR = \frac{FP}{N}$$

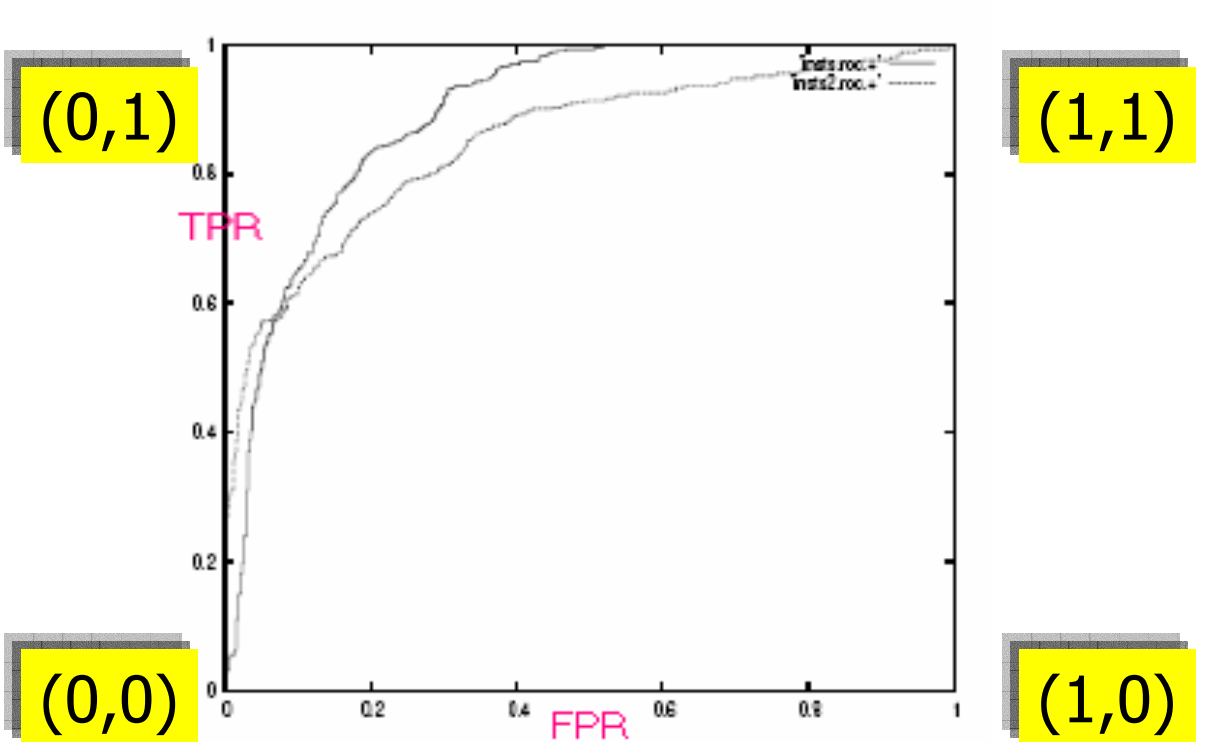
$$Precision = \frac{TP}{TP + FP},\ Accuracy = \frac{TP + TN}{P + N}$$

$$Sensitivity = Recall,\ Specificity = 1 - FPR$$

ROC curve

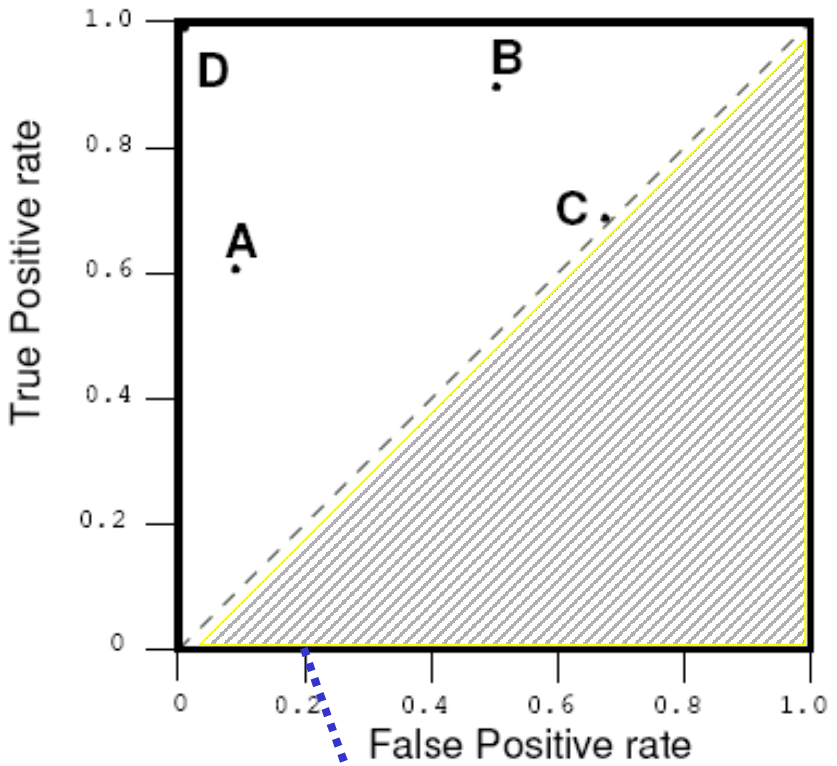
- Y axis: TPR
  X axis: FPR





Compare ROC curve

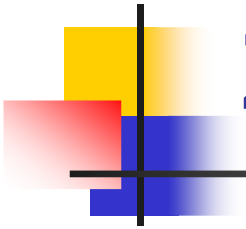
TP	FP
FN	TN



- (0,0) Numbers of P =0,
  → No FP error, No TP
- (0,1) perfect
  → D classifiers
- Northwest location is better.
- Near x axis and on the left side
  → Conservative
  e.g. A vs. B
- Near upper right-hand side
  → Liberal

y=x (?)

Lower Right ? Triangle

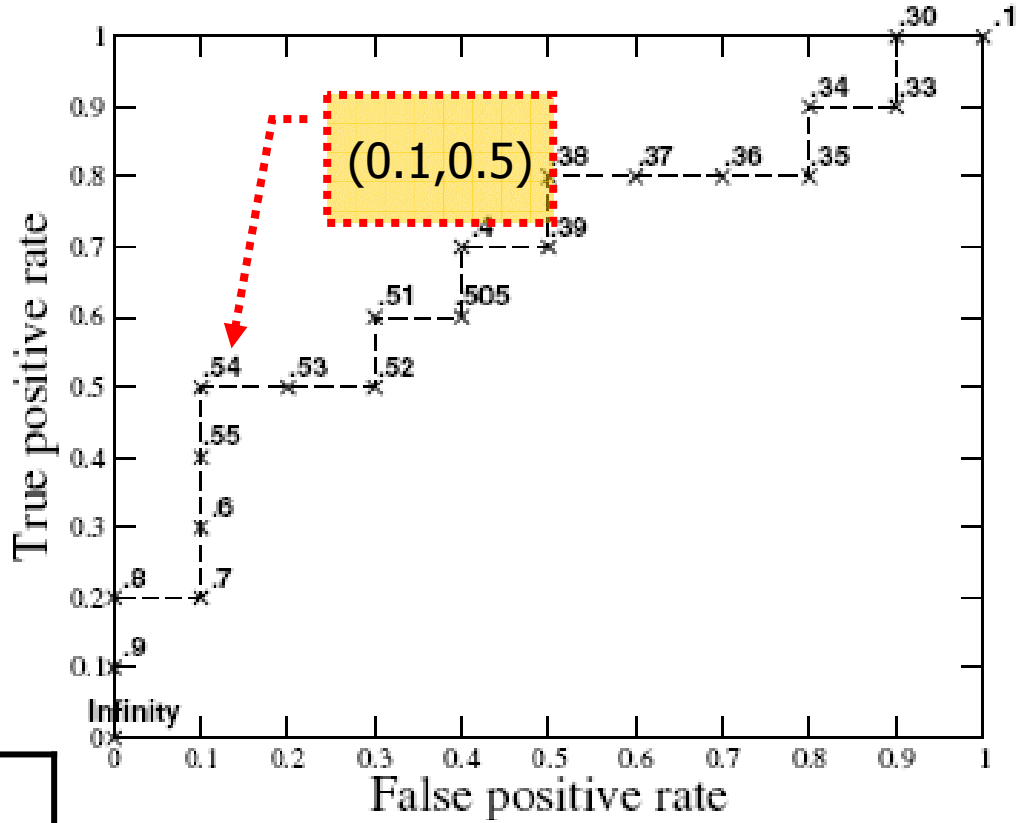


2. Create an ROC curve

- A ranking or scoring classifier can be used with a threshold to produce a binary classifier.
- If the classifier output is above the threshold, the classifier produces a Y, else a N.

Use thresholds to create ROC curve

Inst#	Class	Score	Inst#	Class	Score
1	p	.9	11	p	.4
2	p	.8	12	n	.39
3	n	.7	13	p	.38
4	p	.6	14	n	.37
5	p	.55	15	n	.36
6	p	.54	16	n	.35
7	n	.53	17	p	.34
8	n	.52	18	n	.33
9	p	.51	19	p	.30
10	n	.505	20	n	.1



If threshold =0.54 →
Numbers of Score ≧0.54 → 6

5	1	6
5	9	14
10	10	20

$$x : \frac{1}{10} = 0.1$$

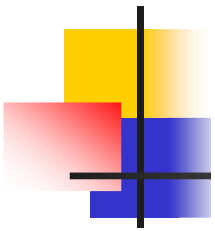
$$y : \frac{5}{10} = 0.5$$

f(i) : the probabilistic classifier's estimate that instance i is positive;
min and max, the smallest and largest values returned by f;
increment : the smallest difference between any two f values.

```
1:  for t = min to max by increment do
2:      FP ⇐ 0
3:      TP ⇐ 0
4:      for i ∈ L do
5:          if f(i) ≥ t then
6:              if i is a positive example then
7:                  TP ⇐ TP + 1
8:              else
9:                  FP ⇐ FP + 1
10:     Add point (FP/N, TP/P) to ROC curve
11: end
```

L Inputs: the set of test instances;

Conceptual Algorithm

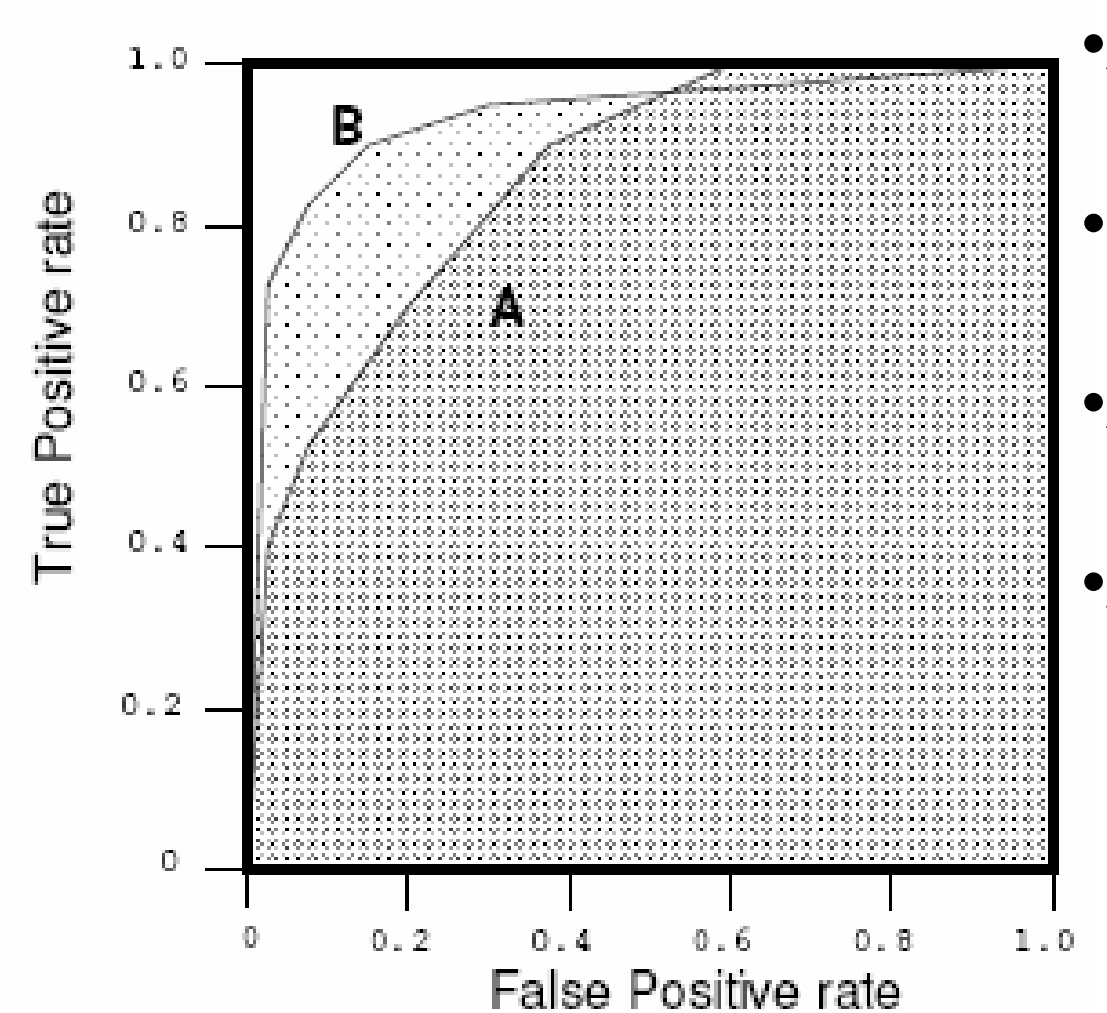


```
1:  L_sorted ⇐ L sorted decreasing by f scores
2:  FP ⇐ 0
3:  TP ⇐ 0
4:  R ⇐ ⟨⟩
5:  f_prev ⇐ −∞
6:  for i ∈ L_sorted do
7:     if f(i) ≠ f_prev then
8:        ADD_POINT(((FP/N, TP/P), R)
9:        f_prev ⇐ f(i)
10:    if i is a positive example then
11:       TP ⇐ TP + 1
12:    else
13:       FP ⇐ FP + 1
14: ADD_POINT(((FP/N, TP/P), R)
15: end
```

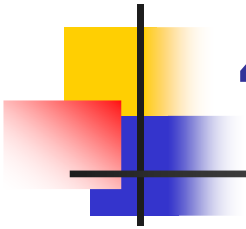
```
1: subroutine ADD_POINT(P, R)
2: push P onto R
3: end subroutine
```

Practical Algorithm

3. **A**rea **U**nder an ROC **C**urve (AUC)

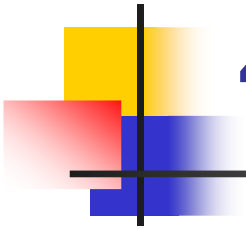


- AUC (Bradley, 1997)
- Wilcoxon test of ranks
- Area : Classifier B > A
- Average performance → B > A



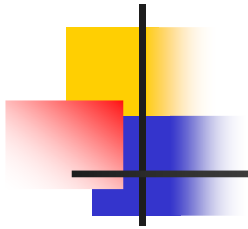
4. R demo – ROCR package

- package : ROCR
- plot ROC curve
- plot SVM vs. Neural Network



4. References

1. Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognition, 30 (7), 1145-1159.
2. Fawcett, T. (2003) ROC Graphs: Notes and Practical Considerations for Data Mining Researchers, HP Laboratories technical report.
3. Witten, I.H. and Frank, E. (2005) Data Mining: Practical Machine Learning Tools and Techniques, Second Edition, Morgan Kaufmann.
4. The magnificent ROC: http://www.anaesthetist.com/mnm/stats/roc/



THANKS

Q & A

Web: http://web.ydu.edu.tw/~alan9956/
Email: alan9956@webmail.ydu.edu.tw