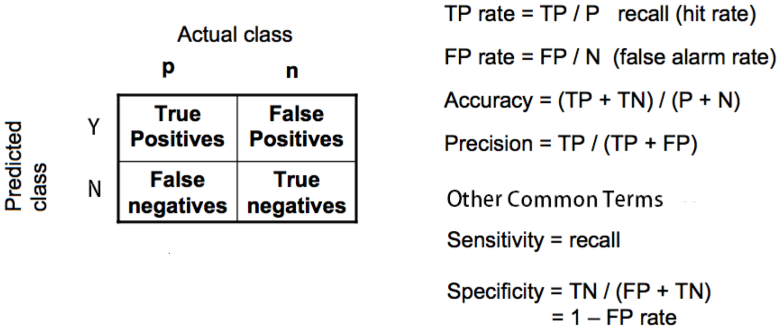
ROC Curves / Profit Curves

Clayton W. Schupp, Galvanize

Spring 2015

Confusion Matrix



		Actual class	
		p	n
Predicted class	Y	True Positives	False Positives
	N	False negatives	True negatives

TP rate = TP / P   recall (hit rate)

FP rate = FP / N (false alarm rate)

Accuracy = (TP + TN) / (P + N)

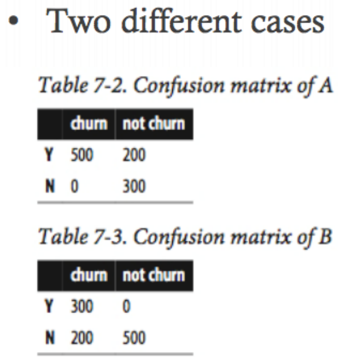
Precision = TP / (TP + FP)

Other Common Terms

Sensitivity = recall

Specificity = TN / (FP + TN)
= 1 – FP rate

Problems with Unbalanced Classes



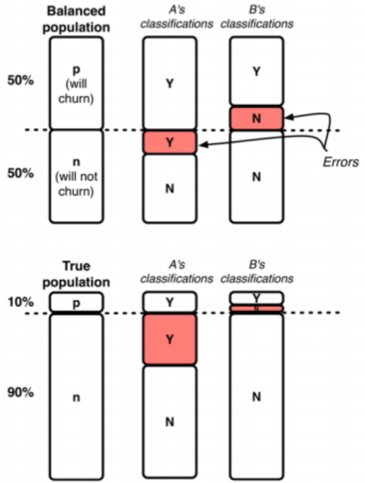
- Two different cases

Table 7-2. Confusion matrix of A

	churn	not churn
Y	500	200
N	0	300

Table 7-3. Confusion matrix of B

	churn	not churn
Y	300	0
N	200	500



Building the ROC Curve

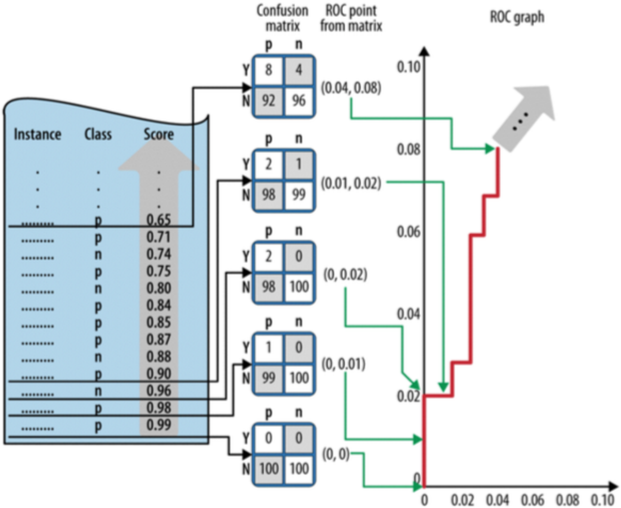
For a given model f, each threshold value T gives a point on the ROC Curve

Model score is the probability of class membership

1. T = minimum score
2. TP=0, FP=0
3. For each observation, i:
    - If $i > T \longrightarrow$ increment TP
    - else $\longrightarrow$ increment FP
4. Add point (FP/N, TP/P) to ROC Graph

Increment T from min-score to max-score, repeating steps 1-4

Building the ROC Curve



Instance	Class	Score
.	.	.
.	.	.
.	.	.
....	p	0.65
....	p	0.71
....	n	0.74
....	p	0.75
....	n	0.80
....	p	0.84
....	p	0.85
....	p	0.87
....	n	0.88
....	p	0.90
....	n	0.96
....	p	0.98
....	p	0.99

Confusion matrix

	p	n
Y	8	4
N	92	96

	p	n
Y	2	1
N	98	99

	p	n
Y	2	0
N	98	100

	p	n
Y	1	0
N	99	100

	p	n
Y	0	0
N	100	100

ROC point from matrix

(0.04, 0.08)

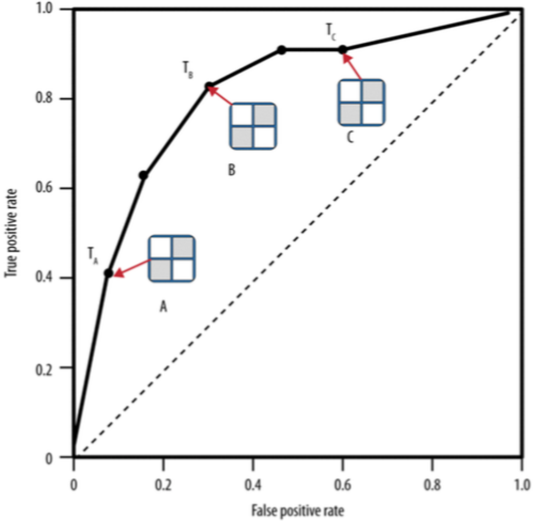
(0.01, 0.02)

(0, 0.02)

(0, 0.01)

(0, 0)

Sample ROC Curve

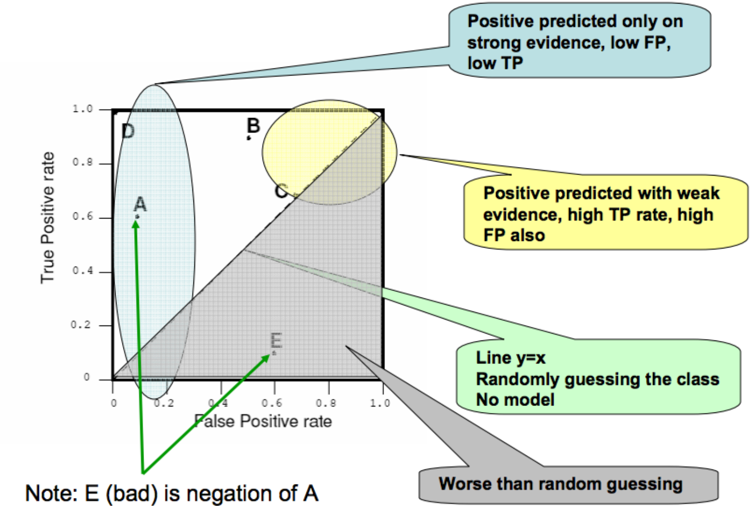


Choosing Between Models

How do we go about choosing a model based on the ROC curve?

- Depends on the goal of the model
    - Screening Test vs. Diagnostic Test
- We can examine the regions of the ROC curve based on desired result

Regions of the ROC Curve



Positive predicted only on strong evidence, low FP, low TP

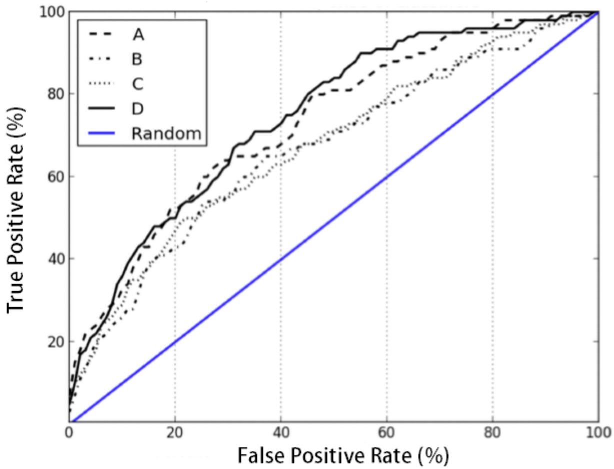
Positive predicted with weak evidence, high TP rate, high FP also

Line y=x
Randomly guessing the class
No model

Worse than random guessing

Note: E (bad) is negation of A

ROC Curve for Multiple Classifiers

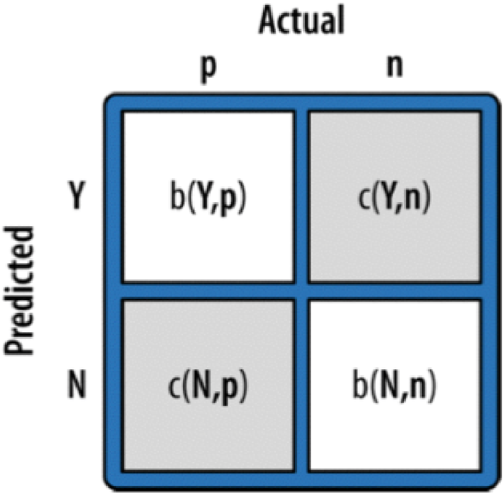


Cost-Benefit Information

- ROC Curves alone assume equal cost of misclassification
- Different kinds of errors have different costs associated
- Correct classifications could also have different benefits

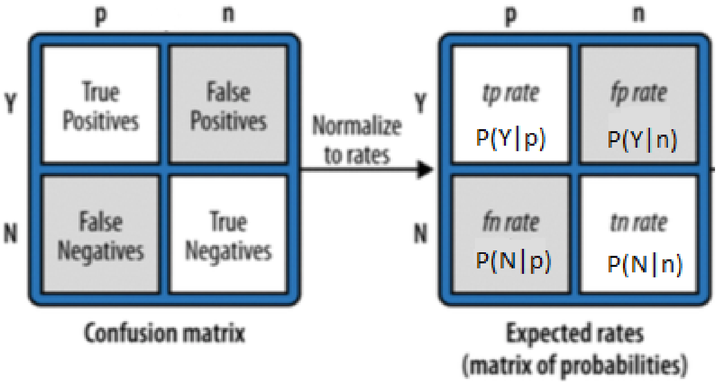
Profit Curves allow us to compare models and select the one that will maximize profit for a specified cost-benefit

Cost-Benefit Matrix



		Actual	
		p	n
Predicted	Y	b(Y,p)	c(Y,n)
	N	c(N,p)	b(N,n)

Normalize Confusion Matrix to Rates



	p	n
Y	True Positives	False Positives
N	False Negatives	True Negatives

Confusion matrix

Normalize to rates

	p	n
Y	*tp rate* P(Y|p)	*fp rate* P(Y|n)
N	*fn rate* P(N|p)	*tn rate* P(N|n)

Expected rates (matrix of probabilities)

Expected Profit

By combining information from the Confusion Matrix and the Cost-Benefit Matrix, we can calculate the Expected Profit:

$$\begin{aligned}
E[Profit] &= P(Y,p)\cdot b(Y,p) + P(Y,n)\cdot c(Y,n) + \\
&= P(N,p)\cdot c(N,p) + P(N,n)\cdot b(N,n) \\
&= P(Y|p)\cdot P(p)\cdot b(Y,p) + P(Y|n)\cdot P(n)\cdot c(Y,n) + \\
&= P(N|p)\cdot P(p)\cdot c(N,p) + P(N|n)\cdot P(n)\cdot b(N,n) \\
&= P(p)\cdot [P(Y|p)\cdot b(Y,p) + P(N|p)\cdot c(N,p)] + \\
&= P(n)\cdot [P(Y|n)\cdot c(Y,n) + P(N|n)\cdot b(N,n)]
\end{aligned}$$

Building the Profit Curve

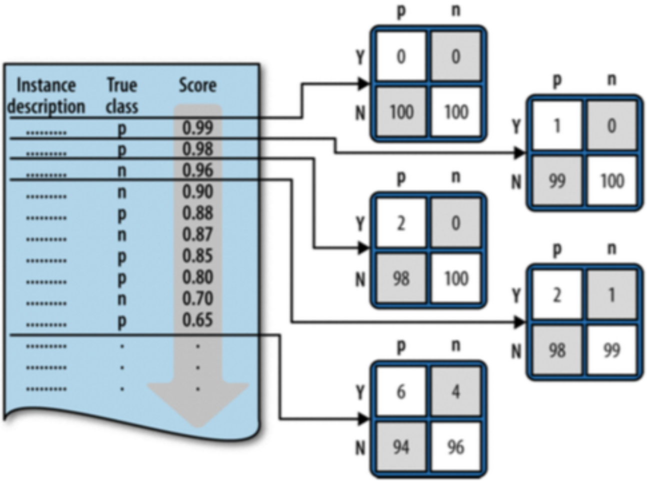
Similar to building ROC Curve. For a given model f, each threshold value T gives a point on the Profit Curve

Model score is the probability of class membership

1. T $=$ maximum score
2. Using the confusion matrix and cost-benefit matrix, calculate $E[Profit]$
3. For each observation, i:
    - If $i > T \longrightarrow$ increment TP
    - else $\longrightarrow$ increment FP
4. Add point ($E[Profit]$, % Test Instances) to Profit Graph

Increment T from max-score to min-score, repeating steps 1-4

Building the Profit Curve



Instance description	True class	Score
....	p	0.99
....	p	0.98
....	n	0.96
....	n	0.90
....	p	0.88
....	n	0.87
....	p	0.85
....	p	0.80
....	n	0.70
....	p	0.65
....	.	.
....	.	.
....	.	.

	p	n
Y	0	0
N	100	100

	p	n
Y	1	0
N	99	100

	p	n
Y	2	0
N	98	100

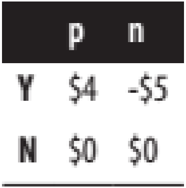
	p	n
Y	2	1
N	98	99

	p	n
Y	6	4
N	94	96

Example: Building the Profit Curve

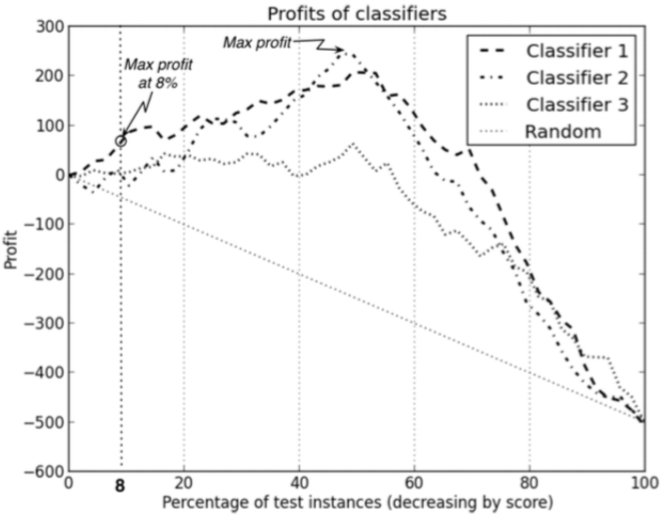
Let's assume our profit margin is small: each offer costs $5 to make and market and each accepted offer earns $9, for a profit of $4.

The cost matrix is:



	p	n
Y	$4	-$5
N	$0	$0

Profit Curves for Multiple Classifiers



Profit Curves Conditions

Critical Condition underlying the profit calculations

- Class Priors: the proportion of positive and negative instances in the target population
- Costs/Benefits: sensitive to relative levels of costs and benefits

sklearn models

Want to point out the following:

- Because of 0/1 classification, confusion matrix lists negatives then positives
- The productive probabilities has two columns, one for each class
  - The first column is for '0' class
  - The second column is for '1' class