Symbols, Patterns and Signals: **Evaluation**

Dr. Iván Palomares Carrascosa, Office 319 Magg's House

Email: i.palomares@bristol.ac.uk

In this lecture...

- **Evaluating classifiers: accuracy**
- **True and False Positive rates**
- **ROC Curve**

Evaluating and tuning classifiers

	Predicted positive	*Predicted negative*	
Positive examples	true positives ***TP***	false negatives ***FN***	***Pos***
Negative examples	false positives ***FP***	true negatives ***TN***	***Neg***
	PPos	***PNeg***	***N***

Normally, classifiers are evaluated by "recording" *N* predictions on a test set in a confusion matrix, and calculating their *accuracy = (TP+TN)/N*

- *error = 1 – accuracy = (FP+FN)/N*

However, this assumes positives and negatives as equally important

- profit of an extra true positive = profit of an extra true negative = +1/N
- cost of an extra false positive = cost of an extra false negative = -1/N

Imbalanced misclassification costs

Medical diagnosis

- the cost of a false negative (patient wrongly classified as healthy) is often much higher than the cost of a false positive (patient wrongly classified as having disease)
- so good performance in predicting disease (positive class) is more important than good performance in predicting healthy condition (negative class)

Imbalanced misclassification costs

Spam email filtering

- the cost of a false positive (non-spam wrongly classified as spam) is often much higher than the cost of a false negative (spam wrongly classified as non-spam)
- so good performance in predicting non-spam (negative class) is more important than good performance in predicting spam (positive class)

True and false positive rates

True positive rate (sensitivity, recall): ***tpr = TP/Pos = TP/(TP+FN)***

- proportion of positives correctly predicted

False positive rate (fall-out): ***fpr = FP/Neg = FP/(FP+TN)***

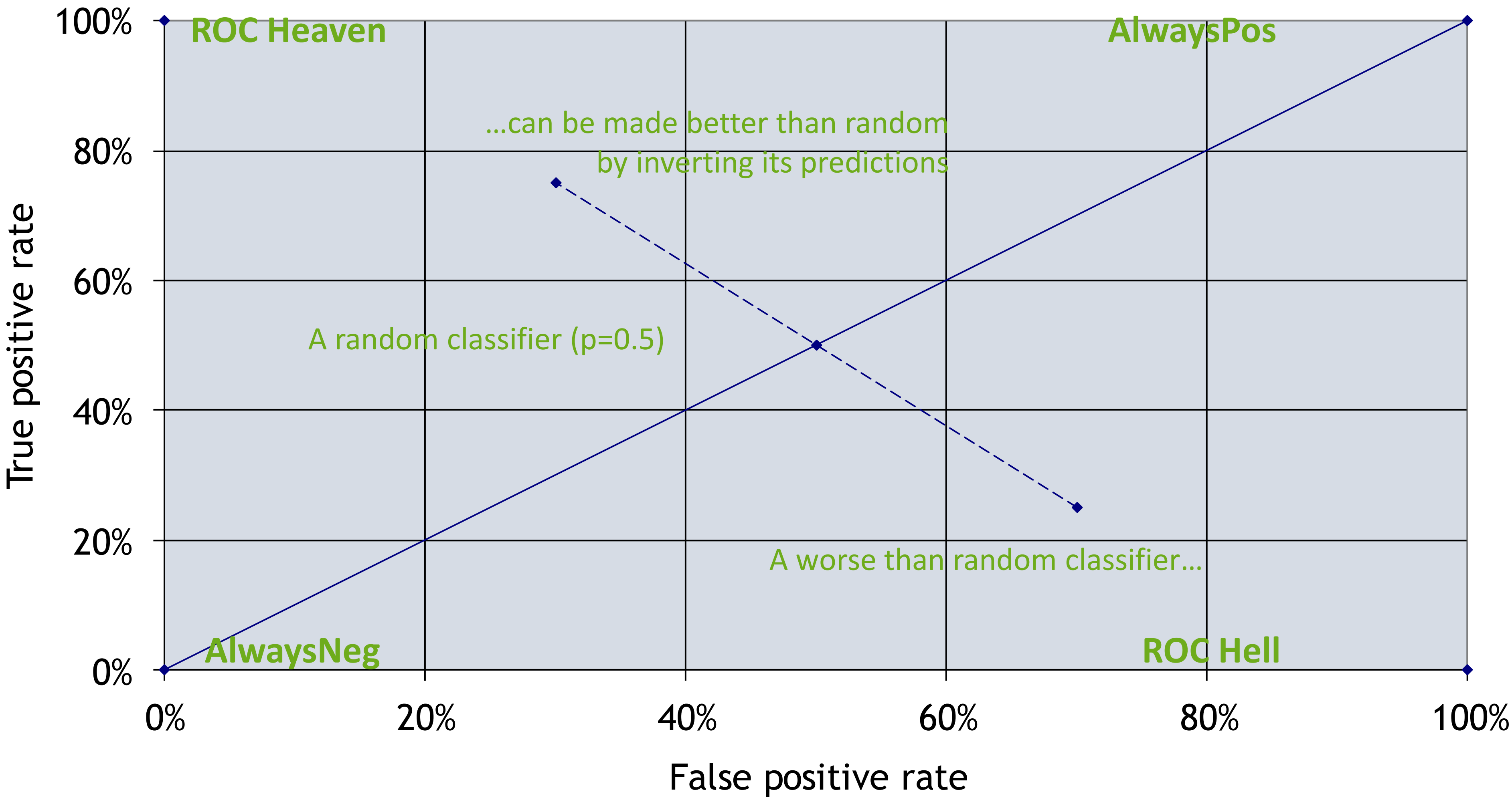
- proportion of negatives incorrectly predicted as positive
- = 1 – true negative rate TN/(FP+TN)

ROC graphs plot true positive rate against false positive rate

- ROC = Receiver Operating Characteristic

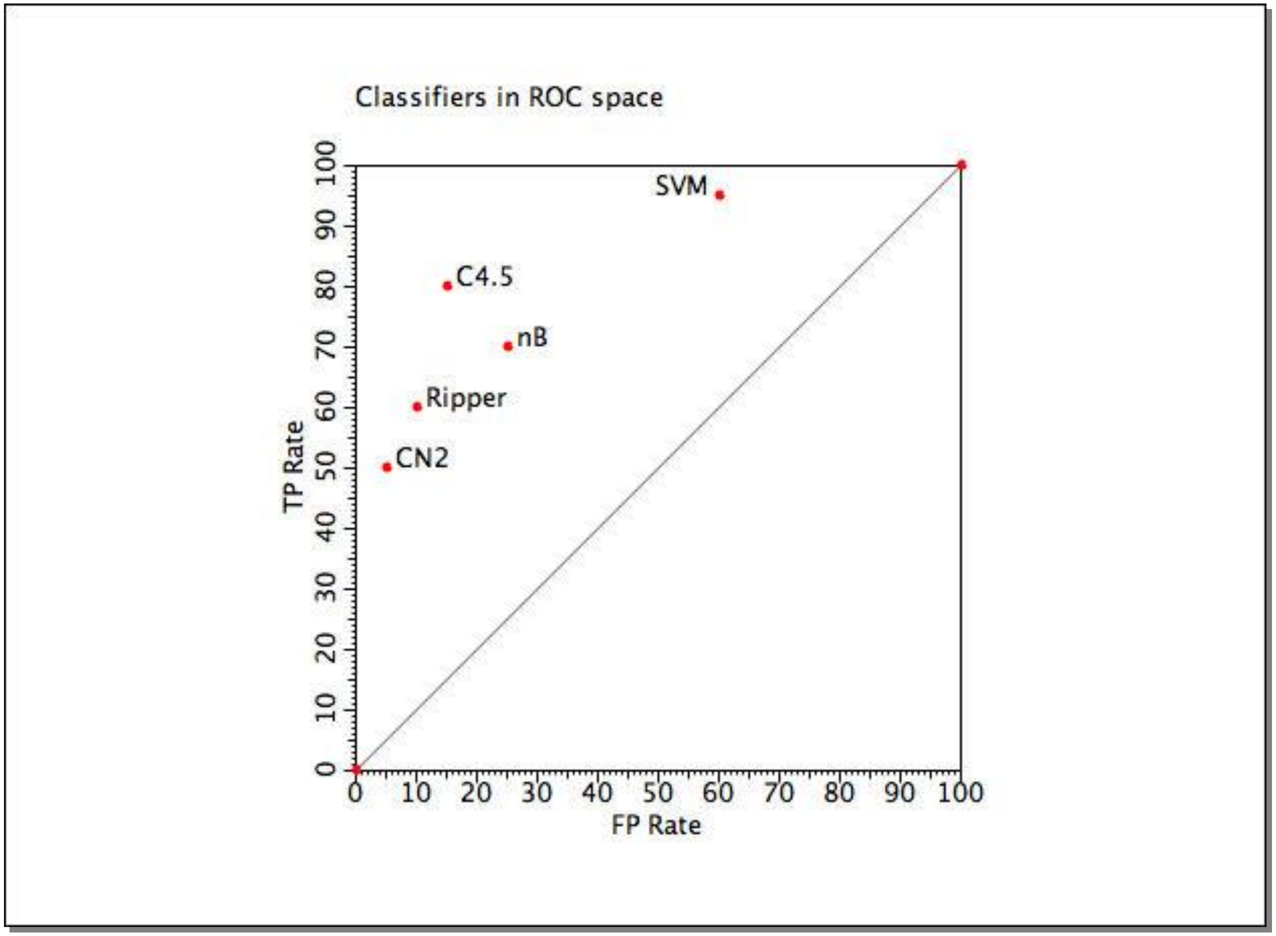
	Predicted positive	*Predicted negative*	
Positive examples	***TP***	***FN***	***Pos***
Negative examples	***FP***	***TN***	***Neg***
	PPos	***PNeg***	***N***

A closer look at ROC space



ROC plot for classifiers

In order to choose the best classifier, we need to take cost factors into account



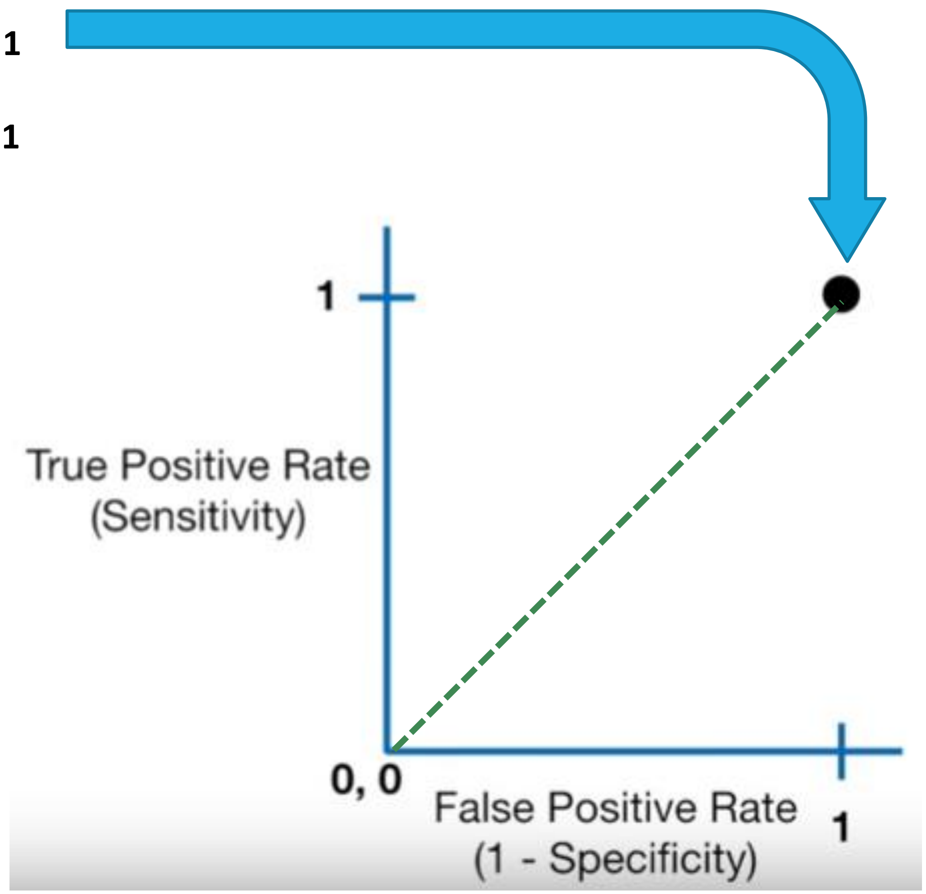
ROC plot produced by ROCon (http://www.cs.bris.ac.uk/Research/MachineLearning/rocon/)

A closer look at ROC space

	P. Diabetic	P. Not Diabetic
Is Diabetic	4	0
Is not Diabetic	4	0

tpr = 4/(4+0) = 1

fpr = 4/(4+0) = 1



A closer look at ROC space

	P. Diabetic	P. Not Diabetic
Is Diabetic	4	0
Is not Diabetic	4	0

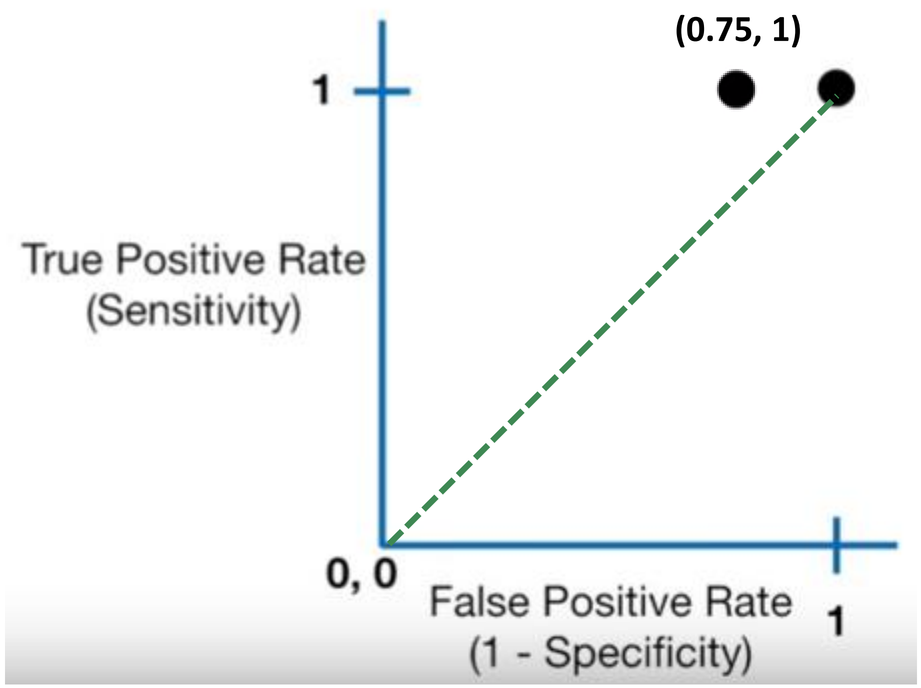
tpr = 4/(4+0) = 1

fpr = 4/(4+0) = 1

	P. Diabetic	P. Not Diabetic
Is Diabetic	4	0
Is not Diabetic	3	1

tpr = 1

fpr = 0.75



A closer look at ROC space

	P. Diabetic	P. Not Diabetic
Is Diabetic	4	0
Is not Diabetic	4	0

tpr = 4/(4+0) = 1

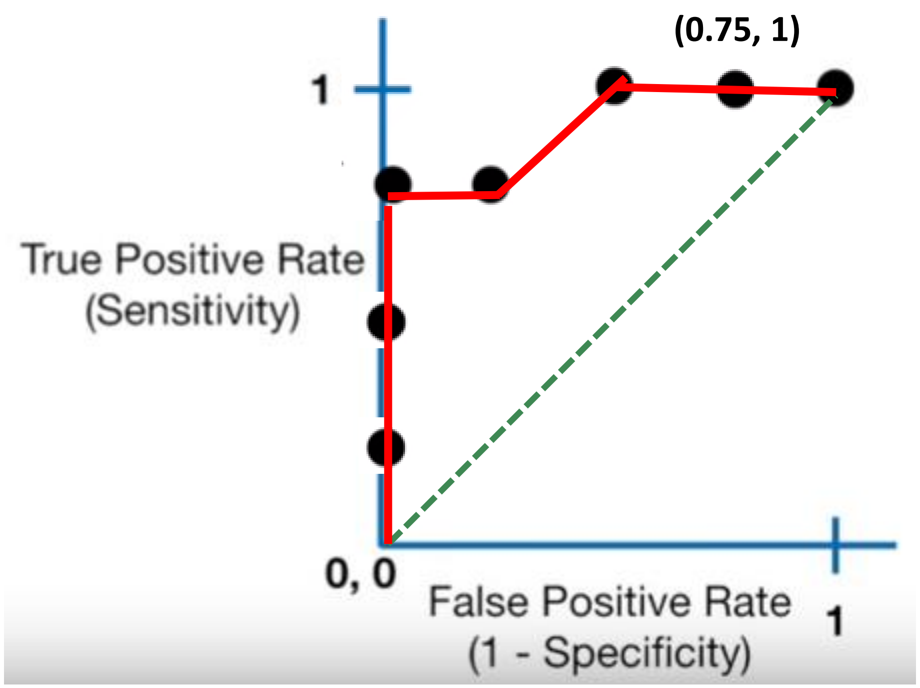
fpr = 4/(4+0) = 1

	P. Diabetic	P. Not Diabetic
Is Diabetic	4	0
Is not Diabetic	3	1

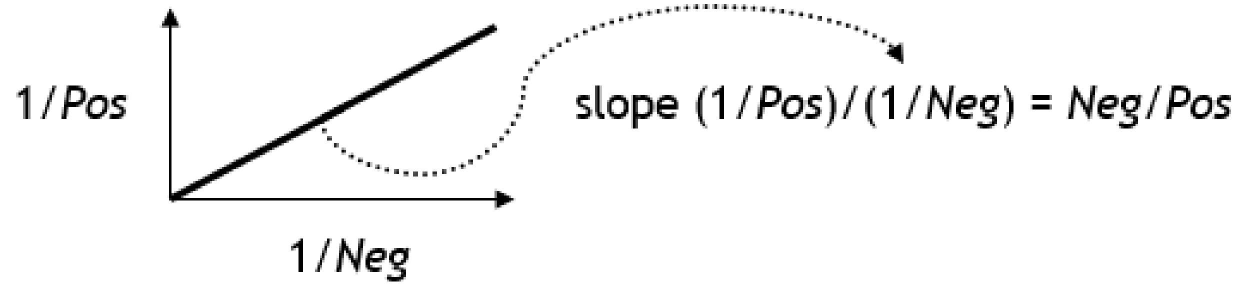
tpr = 1

fpr = 0.75

The ROC graph summarises all the confusion matrices that each threshold (classifier configuration) produced



Cost ratio



- 1 extra true positive increases **tpr** with **1/Pos**
- 1 extra false positive increases **fpr** with **1/Neg**
- If these are worth the same, then all ROC points on a line with slope **Neg/Pos** are equally good

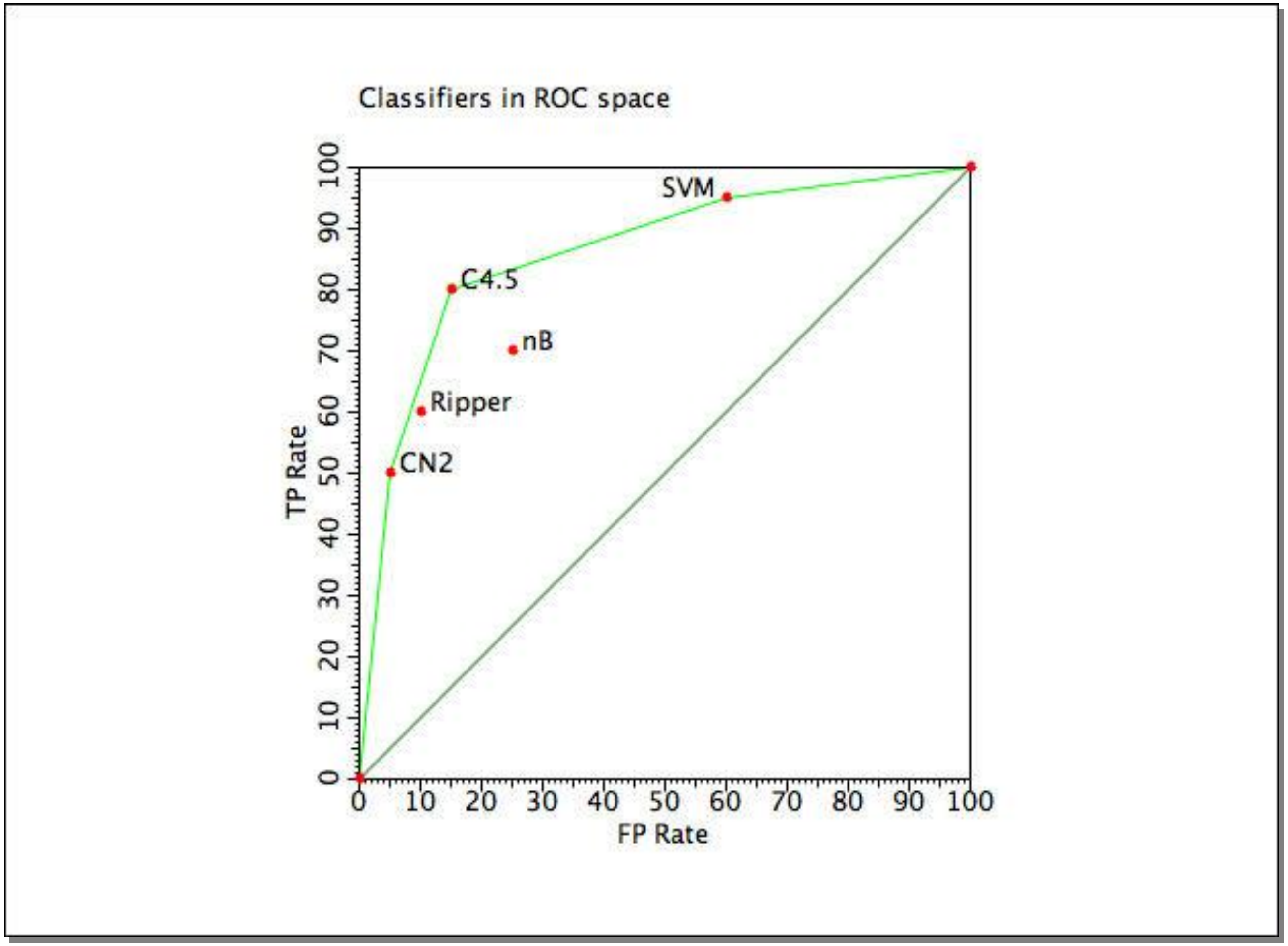
More generally, if *a* extra true positives are worth one extra false positive, and if there are *b* negatives to every positive, then all ROC points on a line with slope *c = a*b* are equally good

c **is called cost ratio and indicates relative importance of tpr and *fpr***

- **c>1**: low **fpr** more important than high **tpr**
- **c<1**: high **tpr** more important than low **fpr**

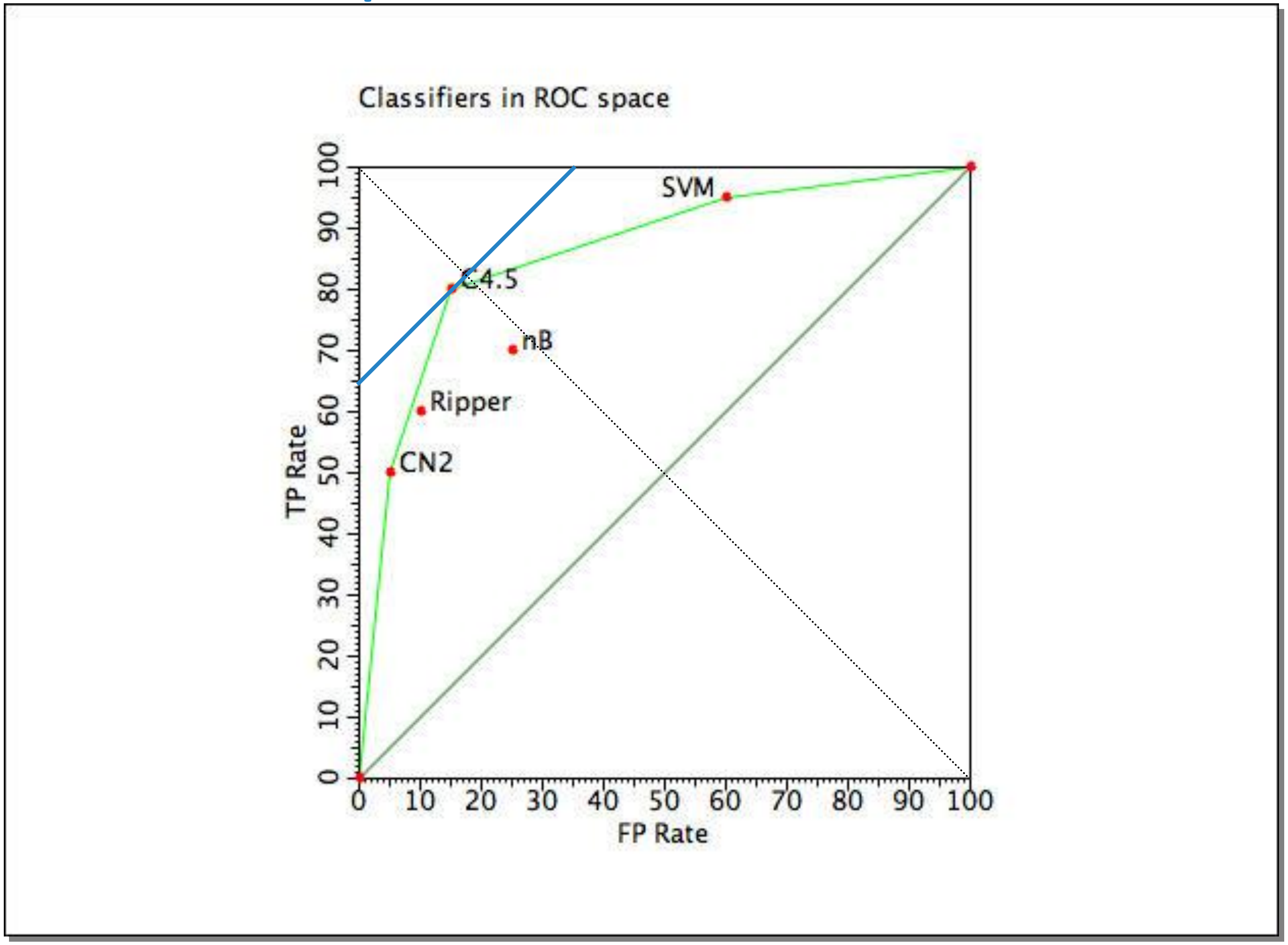
The ROC convex hull

- **The slope of each line segment is the cost ratio under which the two classifiers at the end points are equally good**
  - higher cost ratio ⇒ left one better,
  - lower cost ratio ⇒ right one better
  - <u>classifiers under the convex hull are sub-optimal</u>



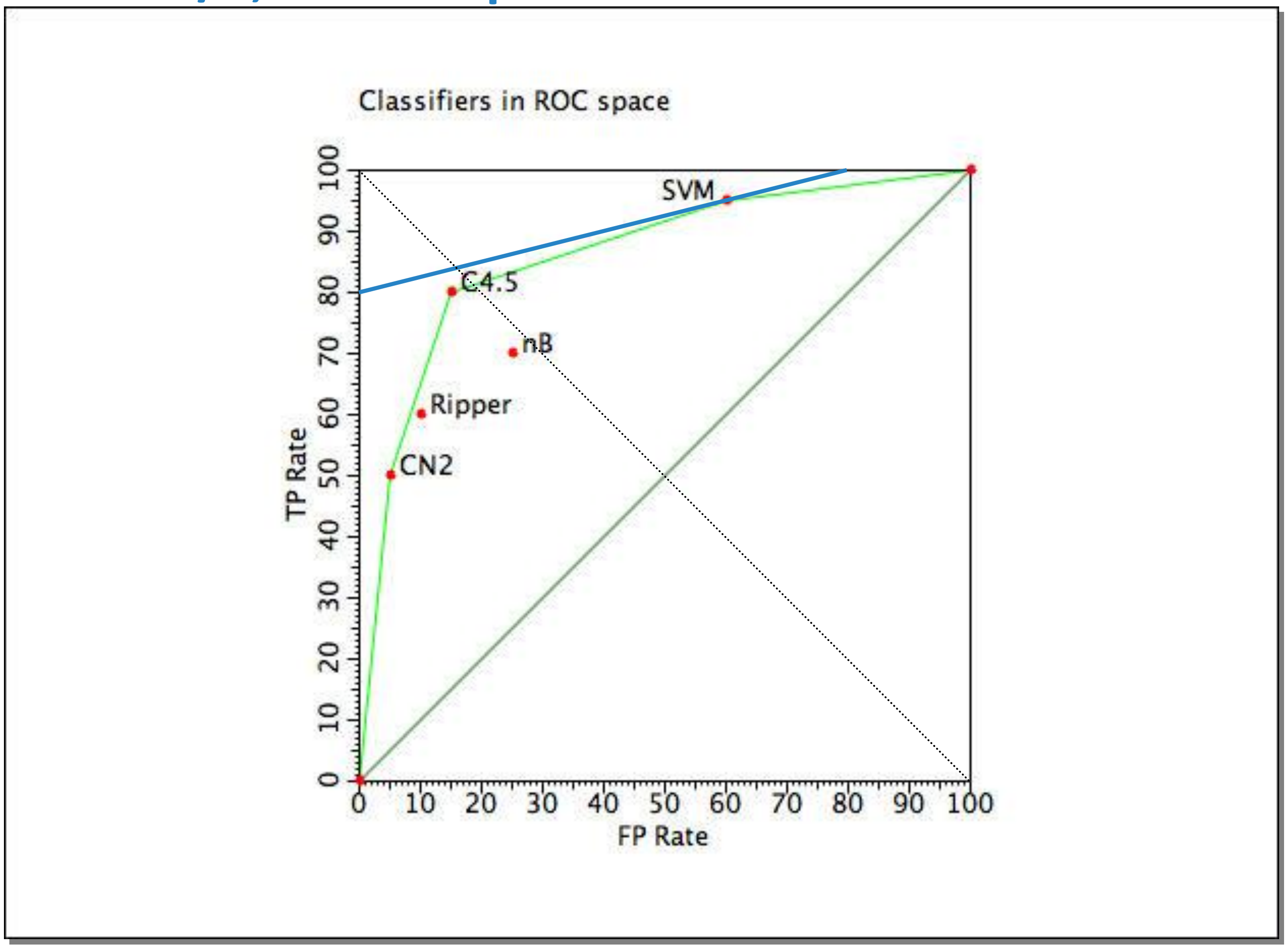
Selecting the optimal classifier

For c=1, C4.5 is optimal



Selecting the optimal classifier

For c=1/4, SVM is optimal



ROC analysis for Bayesian classifiers

Remembering maximum likelihood decision rule:

- **if** $\frac{p(x \mid +)}{p(x \mid -)} \geq 1$ **then** class + **else** class –

We can take class distribution into account by using the maximum a posteriori decision rule:

- **if** $\frac{p(x \mid +)}{p(x \mid -)} \geq \frac{P(-)}{P(+)}$ **then** class + **else** class –

Or we can refrain from setting a specific threshold, i.e. instead of only using the likelihood ratio, let's "play with different thresholds"

- preferable for naïve Bayes classifier, which makes unrealistic independence assumptions, leading to uncalibrated probability estimates

From likelihood ratios to ROC curves

Impractical method:

- consider all possible thresholds on likelihood ratio
  - in fact, only **k+1** for k instances
- construct one confusion matrix for each threshold
- plot in ROC space (each matrix leads to one ***(fpr,tpr)*** point)

Practical method:

- assuming ratio threshold of 1, rank test instances predicted as positive on decreasing likelihood ratio
- starting in **(0,0)**, if the next instance in the ranking is **[+]** move **1/Pos** up, if it is **[-]** move **1/Neg** to the right
  - make diagonal move in case of ties

From likelihood ratios to ROC curves

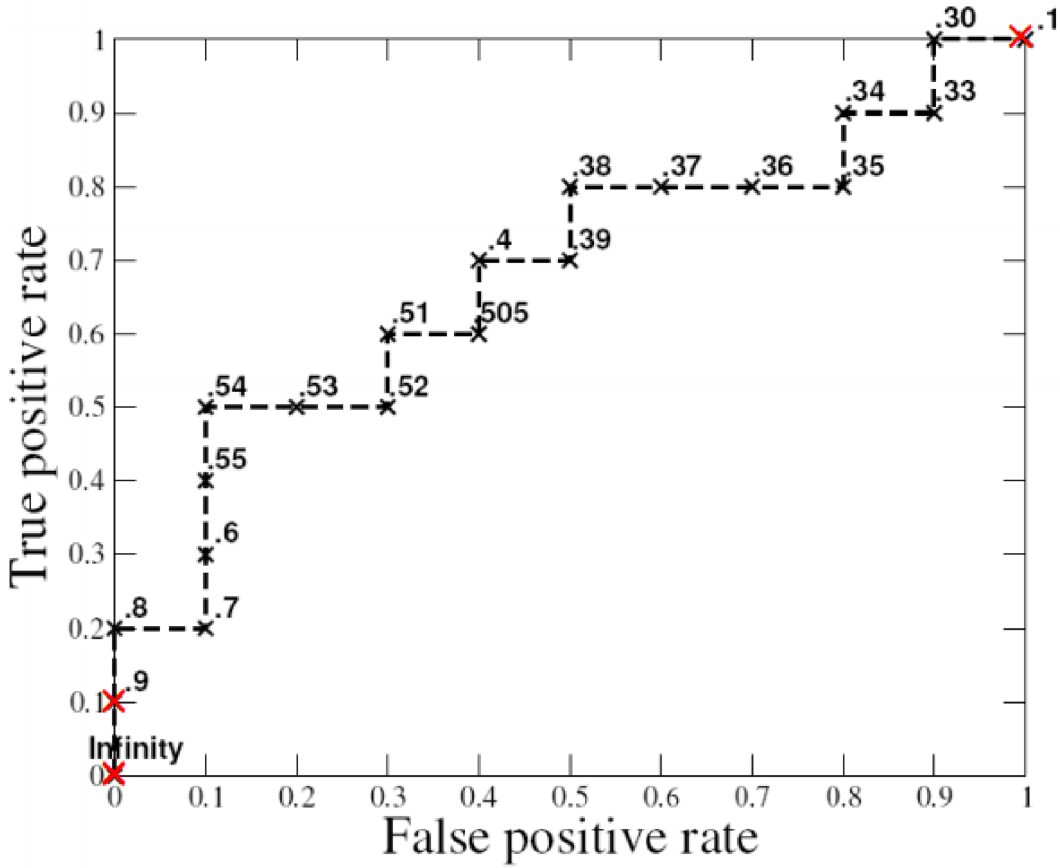
Practical method illustrated

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

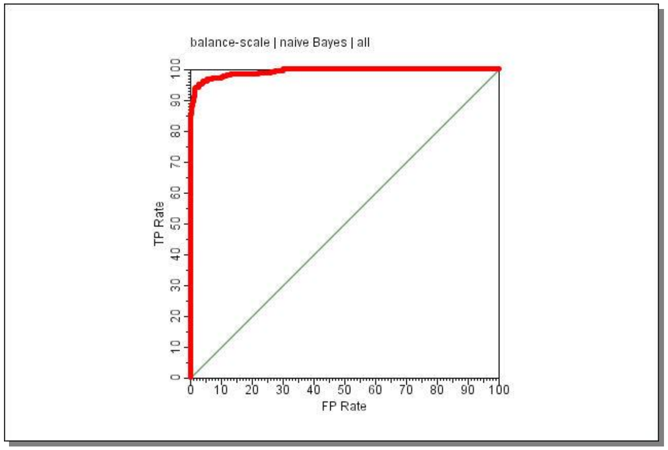
From likelihood ratios to ROC curves

Practical method illustrated

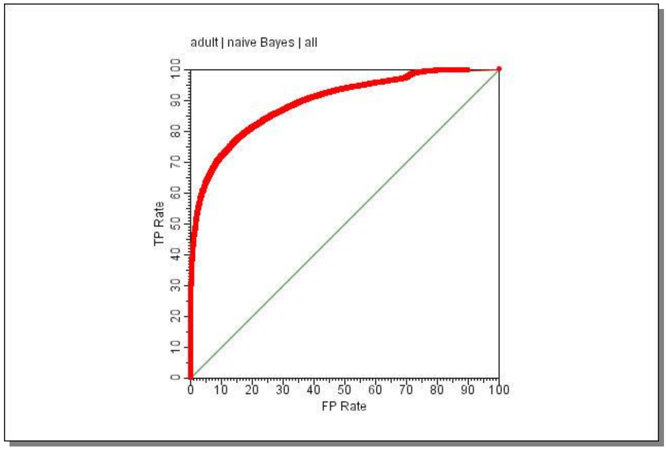
Inst#	Class	Score	Inst#	Class	Score
1	p	.9	11	p	.4
2	p	.8	12	n	.39
3	n	.7	13	p	.38
4	p	.6	14	n	.37
5	p	.55	15	n	.36
6	p	.54	16	n	.35
7	n	.53	17	p	.34
8	n	.52	18	n	.33
9	p	.51	19	p	.30
10	n	.505	20	n	.1



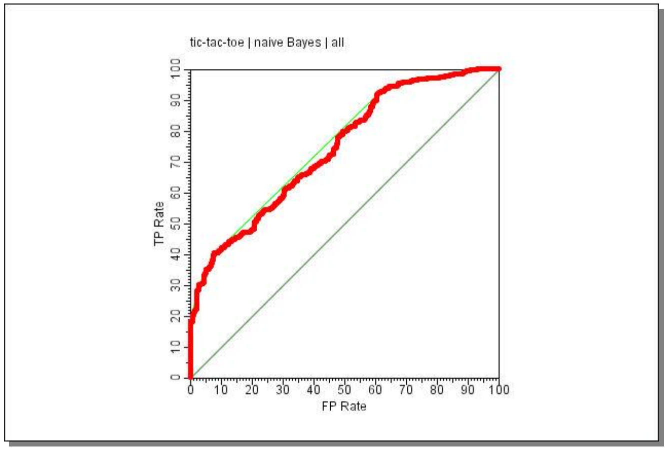
Some example ROC curves



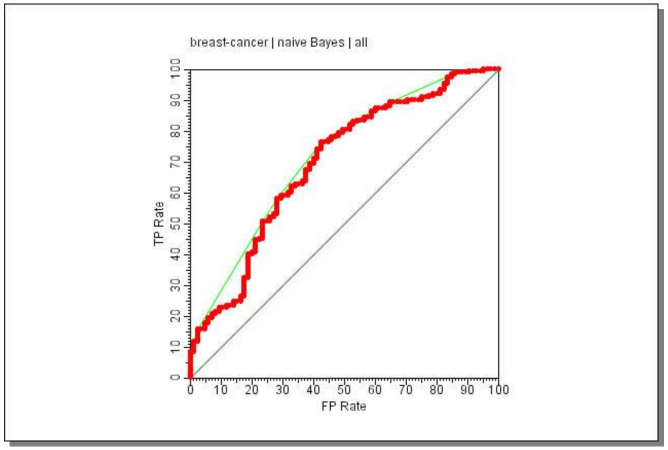
Good separation between classes, convex



Reasonable separation, mostly convex



Fairly poor separation, mostly convex



Poor separation, large and small concavities

The AUC metric for evaluating rankers

The Area Under ROC Curve (AUC) assesses the ranking in terms of separation of the classes: the bigger AUC, the better

- all the **[+]** **“**show up” before the **[-]**$\Leftrightarrow$ **AUC=1**
- random ordering $\Rightarrow$ **AUC=0.5** (but not vice versa)
- all the **[-]** before the **[+]** $\Leftrightarrow$ **AUC=0**

AUC is very frequently used tool to <u>COMPARE CLASSIFIERS</u>