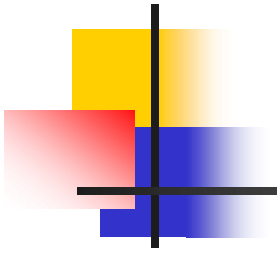
ROC Analysis

Beatrice Lazzerini

Department of Information Engineering
University of Pisa
ITALY



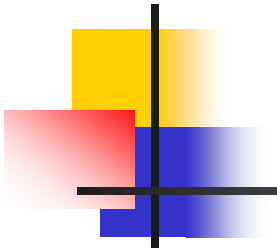
Classifier

- A *classifier* is any function:

$$D : \Re^n \rightarrow \Omega$$

with $\Re^n$ the feature space and Ω the set of class labels.

- A classifier is built from a labeled data set and assigns class labels to objects.
- We consider binary classification
  - a *discrete* classifier outputs a class label,
  - a *continuous* classifier outputs a real value to which different thresholds can be applied to predict class membership.



Classifier performance

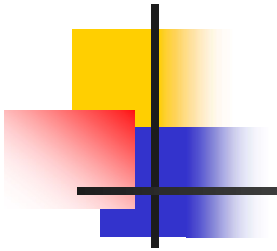
- Classifier comparison is traditionally based on *classification accuracy*:

$$accuracy = \frac{correctly\ classified\ samples}{testing\ samples}$$

Alternatively, $error\ rate = 1 - accuracy$

The use of accuracy assumes that the class prior probabilities are constant and relatively balanced.

It also assumes equal error costs.



- In real-world problems, there is often a large population of normal or uninteresting cases and a small number of unusual or interesting cases (e.g., fraud detection, diagnostic medical tests for rare diseases, etc.), that is, the class distribution is very skewed.

- With skewed class distribution, accuracy does not work. E.g., consider a domain where the classes appear in a 999:1 ratio. A simple rule "*always classify as the maximum likelihood class*" gives 99.9% accuracy (e.g., skews of 10^2 are common in fraud detection).

Let us consider a *two-class* problem (binary classification), e.g., a diagnostic test that tries to determine whether a person has a certain disease:

p = class of positives (e.g., patients with cancer)
n = class of negatives (e.g., patients without cancer)

Let *Y* = predicted positive class
N = predicted negative class

We have four possible outcomes:

True Positive (*TP*):	positive object classified as positive	(*hit*)
False Negative (*FN*):	positive object classified as negative	(*miss*)
True Negative (TN):	negative object classified as negative	(*correct rejection*)
False Positive (*FP*):	negative object classified as positive	(*false alarm*)

Confusion matrix

Let us define an experiment from ***P*** positive instances and ***N*** negative instances. The four outcomes can be represented in a 2x2 *confusion matrix*

		TRUE CLASS p	TRUE CLASS n
PREDICTED CLASS	Y	True Positives	False Positives
	N	False Negatives	True Negatives
Column totals		***P***	***N***

$$accuracy = \frac{TP + TN}{P + N}$$

$$TP\,rate = \frac{TP}{TP + FN} = \frac{TP}{P}$$

$$FP\,rate = \frac{FP}{TN + FP} = \frac{FP}{N}$$

- *true positive rate* (or *hit rate* or *recall*) of a classifier:

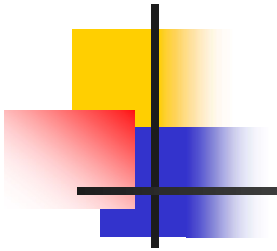
$$TPR = \frac{positives\ correctly\ classified}{total\ \ positives}$$

TPR coincides with *sensitivity*.

- *false positive rate* (or *false alarm rate*) of a classifier:

$$FPR = \frac{negatives\ incorrectly\ classified}{total\ \ negatives}$$

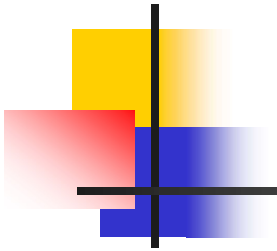
FPR coincides with the complement to 1 of *specificity*.



- Error cost function: *c* (*classification*, *class*)
  - c(Y,n) = cost of a false positive error
  - c(N,p) = cost of a false negative error
- Evaluation by classification accuracy assumes <u>equal error costs</u>:

  c(Y,n) = c(N,p)

- In the real world, this is rarely true, because classifications lead to actions, which have consequences. Actions can be as diverse as denying a credit charge or informing a patient of a cancer diagnosis. Performing an incorrect action can be very costly. The costs of mistakes are rarely equivalent: e.g., not recognizing a cancer is much worse than judging that a healthy patient has cancer. In such cases, accuracy maximization should be replaced with cost minimization.



ROC graphs

- More general classifier comparisons can be made with *Receiver Operating Characteristics* (*ROC*) analysis.
- ROC graphs are a useful technique for visualizing and selecting classifiers based on their performance.
- They are especially useful in the presence of skewed class distribution and unequal classification error costs.

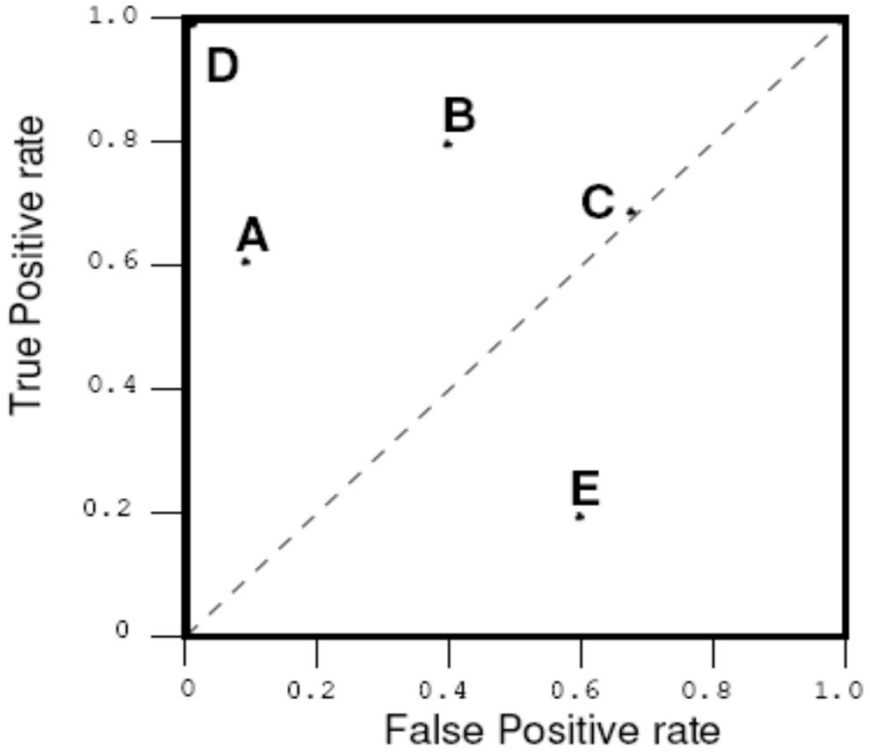
ROC space

ROC graphs are two-dimensional graphs:
TPR is plotted on the *Y* axis *FPR* is plotted on the *X* axis

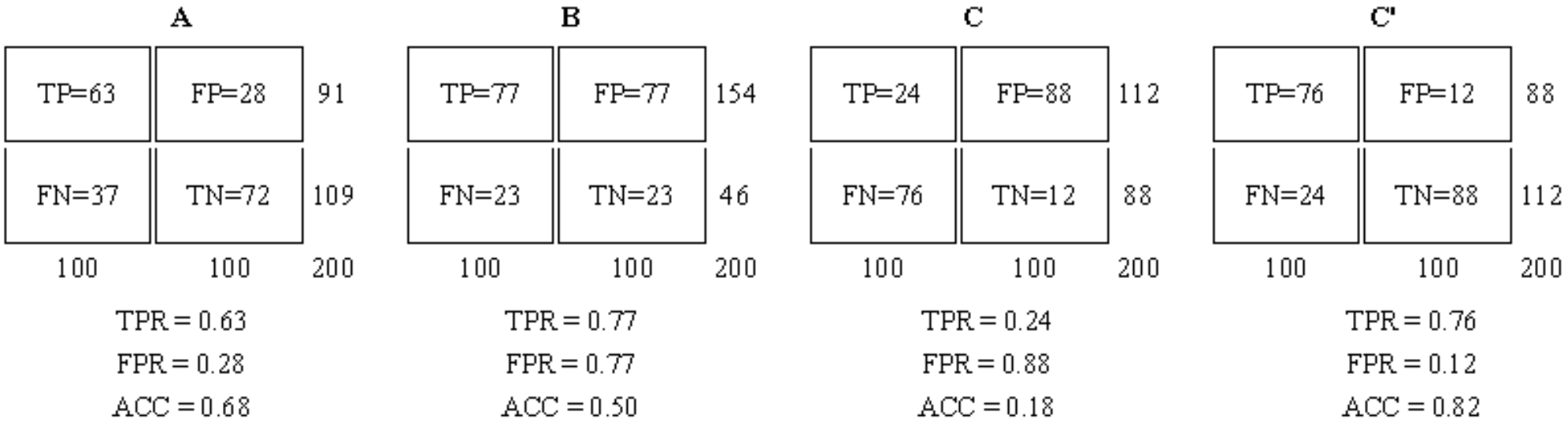
An ROC graph shows relative trade-offs between benefits (true positives) and costs (false positives).

A discrete (binary) classifier (i.e., an instance of a confusion matrix) produces an (FP rate, TP rate) pair corresponding to a single point in the ROC space.

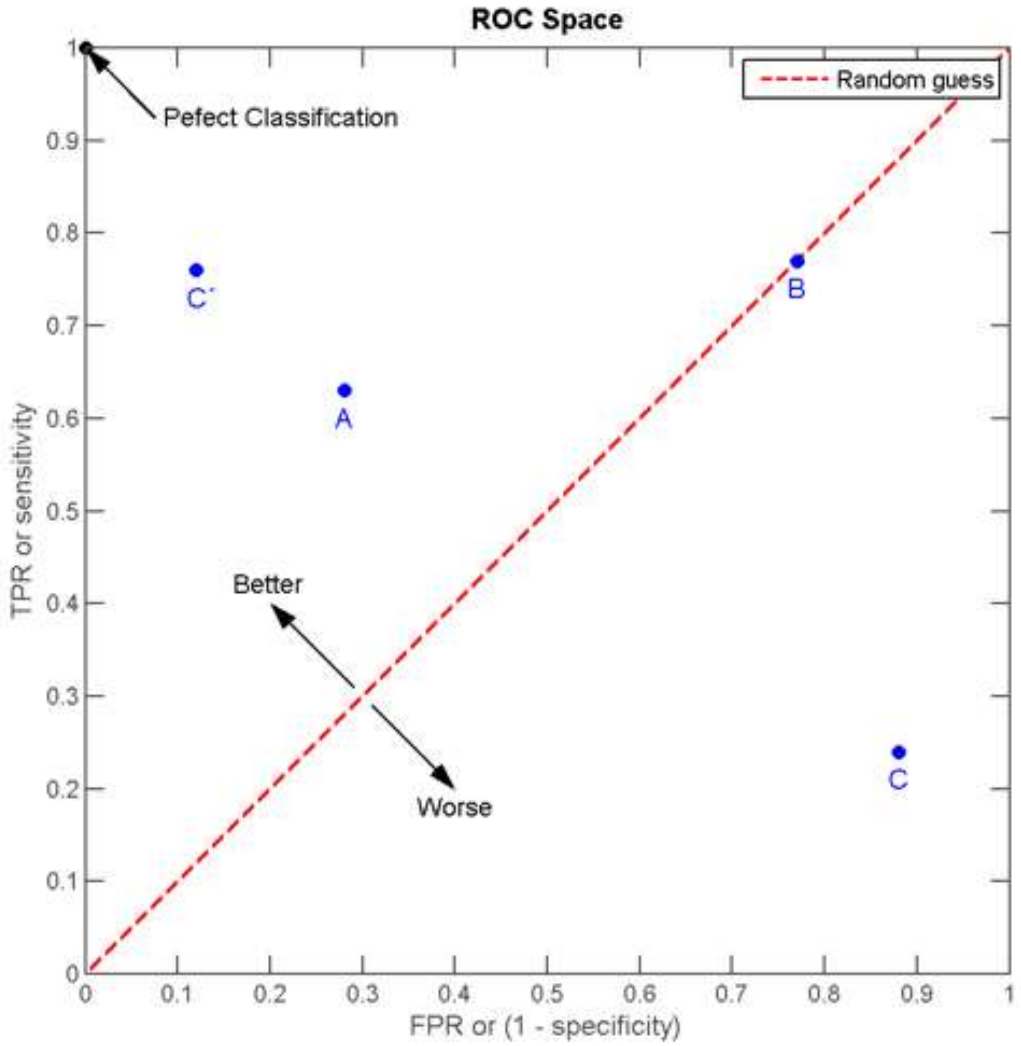
The figure shows an ROC graph showing five discrete classifiers (labeled A to E).



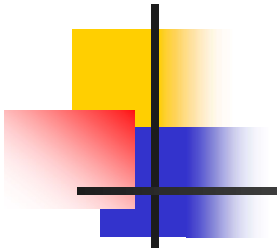
- Let us consider four prediction results from 100 positive and 100 negative instances:



- Let us plot the four results above in the ROC space.



- **A** shows the best predictive power among **A**, **B**, and **C**.
- The accuracy of **B** is 50%, therefore the result of **B** lies on the random guess line.
- **C** shows the worst predictive power among **A**, **B**, and **C**, but simply reversing its decisions leads to a new predictive method **C'** which is even better than **A**.
- The closer the result of a confusion matrix is to the upper left corner, the better the prediction, but the distance from the random guess line in either direction is the best indicator of a method's predictive power.



ROC space

Several points in the ROC space are important:

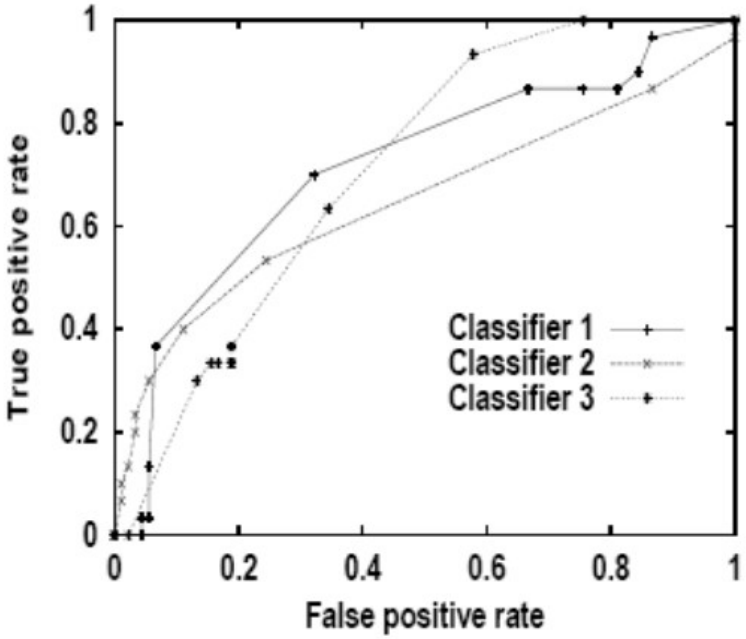
- the lower left point (0,0) represents the strategy of never alarming, i.e., never issuing a positive classification: there is no false positive error but also no true positive,
- the upper right point (1,1) represents the opposite strategy of always alarming, i.e., of unconditionally issuing positive classifications,
- the point (0,1) represents perfect classification,
- the diagonal line y = x represents the strategy of randomly guessing the class (any classifier that appears in the lower right triangle performs worse than random guessing).

Informally, one point in the ROC space is better than another if it is located more north-west (higher TPR, lower FPR, or both).

An ROC graph allows an informal visual comparison of a set of classifiers.

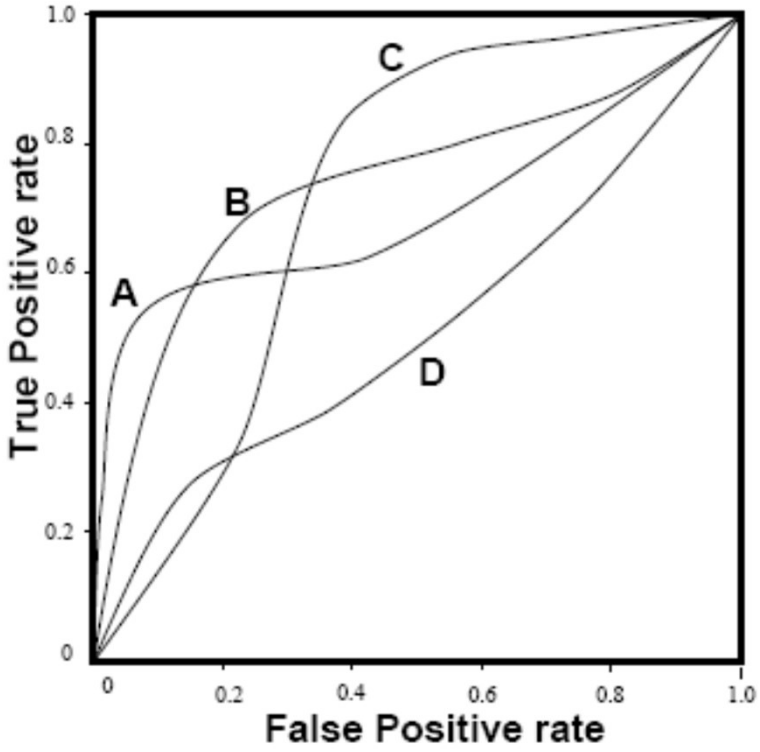
Curves in the ROC space

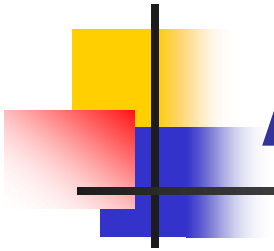
- A continuous classifier is represented by a set of pairs (*FPR*, *TPR*), each related to a specific choice of the threshold, which discriminates negative outputs from positive ones: if the classifier output is above the threshold, the classifier produces *Y*, else *N* (each threshold defines a discrete classifier).
- These pairs give origin to the so-called *ROC curve*.



ROC graphs
of three classifiers

- The ROC curve shows the behavior of a classifier without any reference to the class distribution or the classification error cost, thus decoupling classification performance from these factors.
- Unfortunately, while an ROC graph is a valuable visualization technique, it can help you choose the best global classifier only if there is a classifier that dominates all others across the entire ROC space.
- E.g., curve A is better than curve D because it dominates in all points.





Area Under the Curve (AUC)

- If no classifier dominates all others over the entire ROC space, a frequently adopted method is to choose the classifier with the maximum *area under the curve* (AUC).
- Assuming you are not interested in a specific trade-off between TPR and FPR (i.e., a particular point/region on the ROC curve), the AUC is useful as it aggregates performance across the full range of trade-offs.
- Interpreting the AUC is easy: the higher the AUC, the better, with 0.5 indicating random performance and 1 denoting perfect performance.