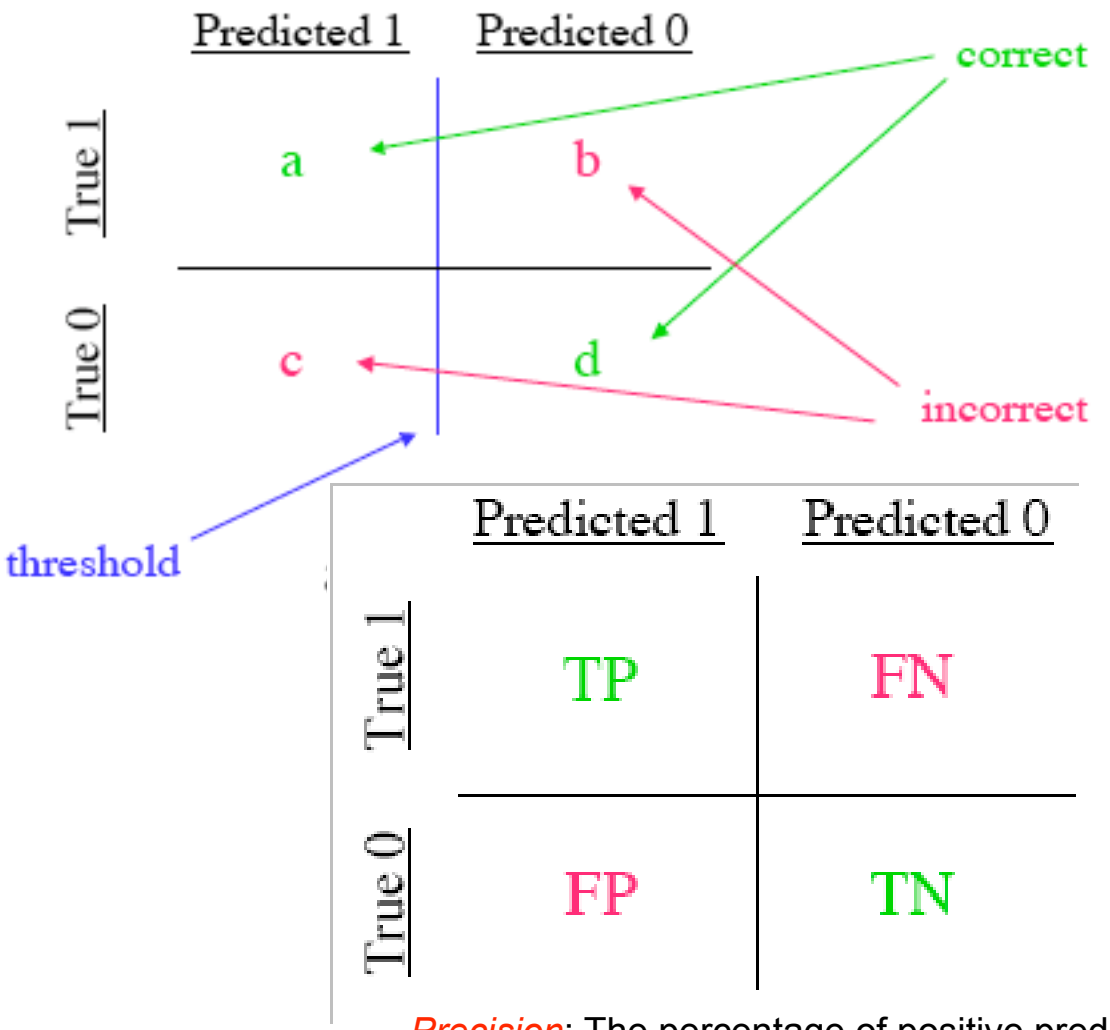
Performance Evaluation



Precision=a/(a+c)
Recall=a/(a+b)
Sensitivity=Recall
Specificity=d/(d+c)
Accuracy=(a+d)/(a+b+c+d)

TPR=*a/(a+b)*
FNR=*b/(a+b)*
FPR=*c/(c+d)*
TNR=d/(c+d)
TPR+FNR=1
FPR+TNR=1

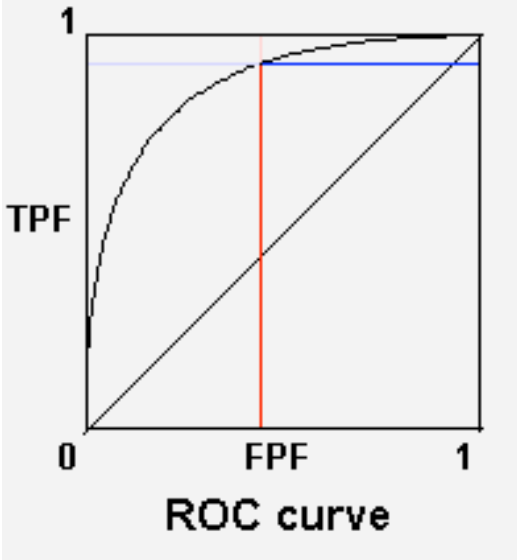
Precision: The percentage of positive predictions that are correct.
Recall: The percentage of positive labeled instances that were predicted as positive.

Performance Evaluation

- ROCs (Receiver Operating Characteristic Curves)
  - They are commonly used for decision-making
  - They summarize the performance of the classifier
  - Using them, it is easy to visualize and analyze the output of the classifier
  - They facilitates the comparison of the performance of different methods
  - True Positives Rate (TPR):
    - Fraction of *positives* (faces), classified as *positives.*
  - True Negative Rate (TNR):
    - Fraction of *negatives* (non-faces), classified as *negatives.*
  - False Positives Rate (FPR):
    - Fraction of *positives*, classified as *negatives.*
  - False Negative Rate (FNR):
    - Fraction of *negatives*, classified as *positives.*

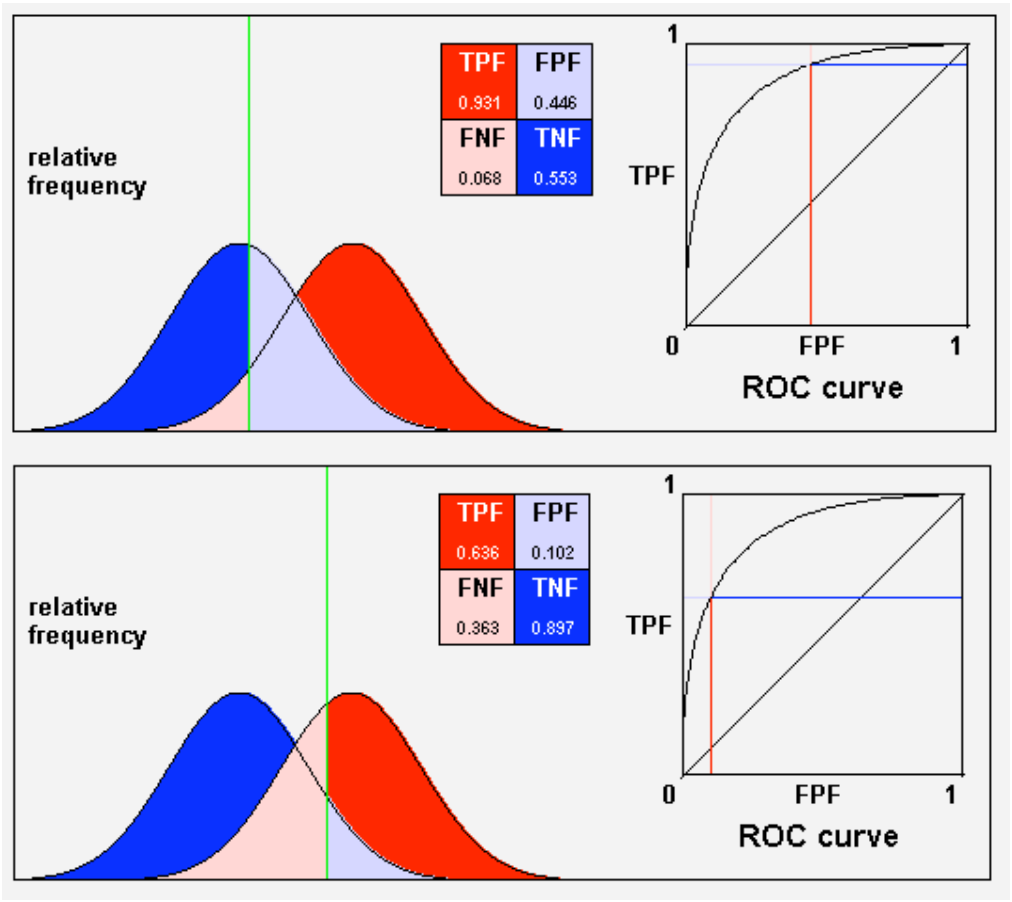
  TPR + FNR = 1

  TNR + FPR = 1



Performance Evaluation

- ROCs
  - A ROC represents the accuracy of the classifier for differentiating the to class.
  - They define how much overlapped are the outputs for the two class (the classifier output distribution for the two class).
  - When different threshold levels are applied to the classifier output, different operation points are obtained
  - The larger the TPR, the larger the FPR



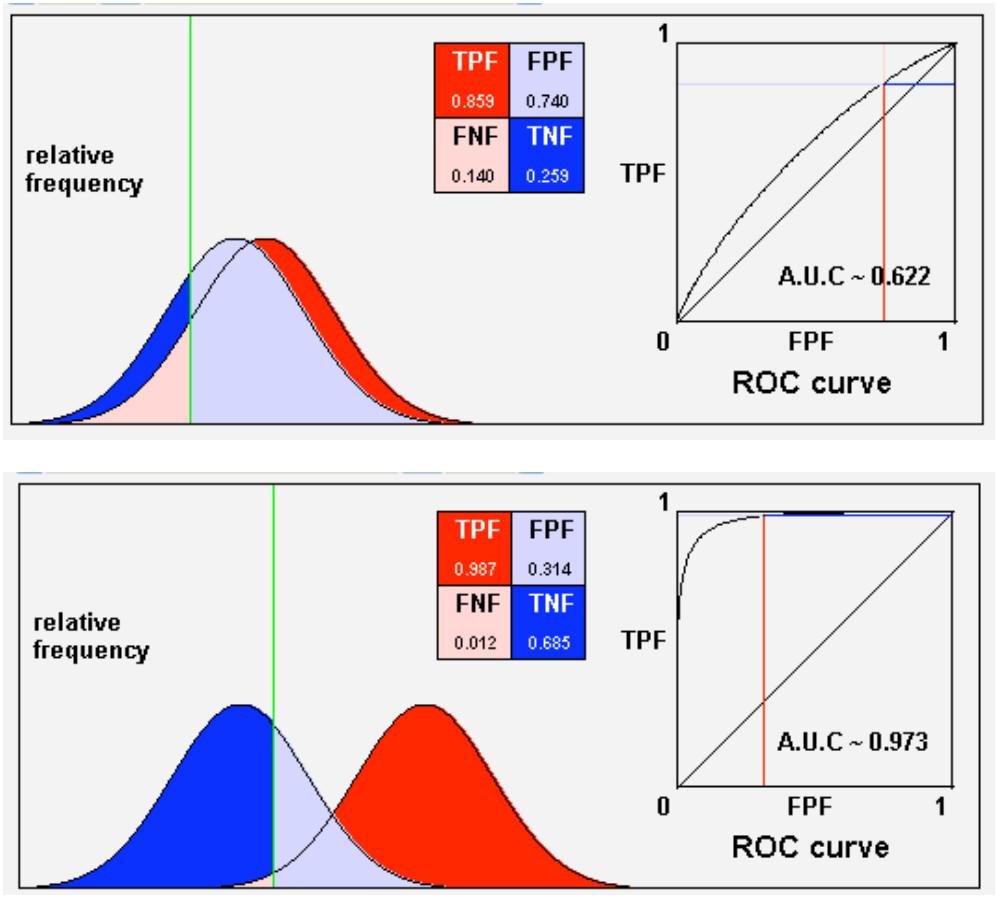
In the example the output of the classifiers each of the class follows a Gaussian distribution for

Images from: http://www.anaesthetist.com/mnm/stats/roc/

Javier Ruiz del Solar and Rodrigo Verschae, EVIC 2004

Performance Evaluation

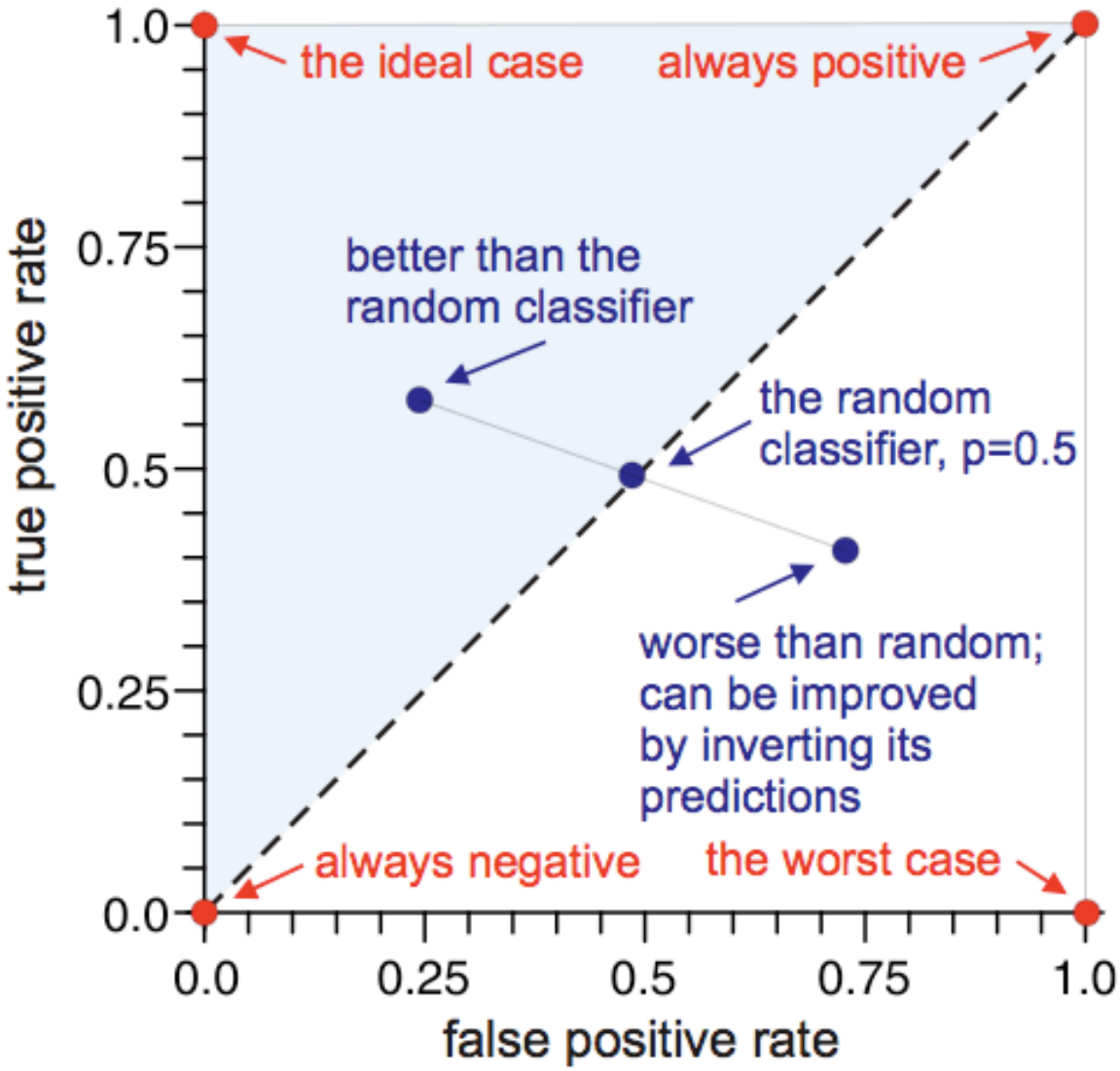
- ROCs
  - The AUC (Area Under the Curve) is used as performance measure in some cases.
  - In detection problems is usual the use of the number of false positives, instead of the FPR.



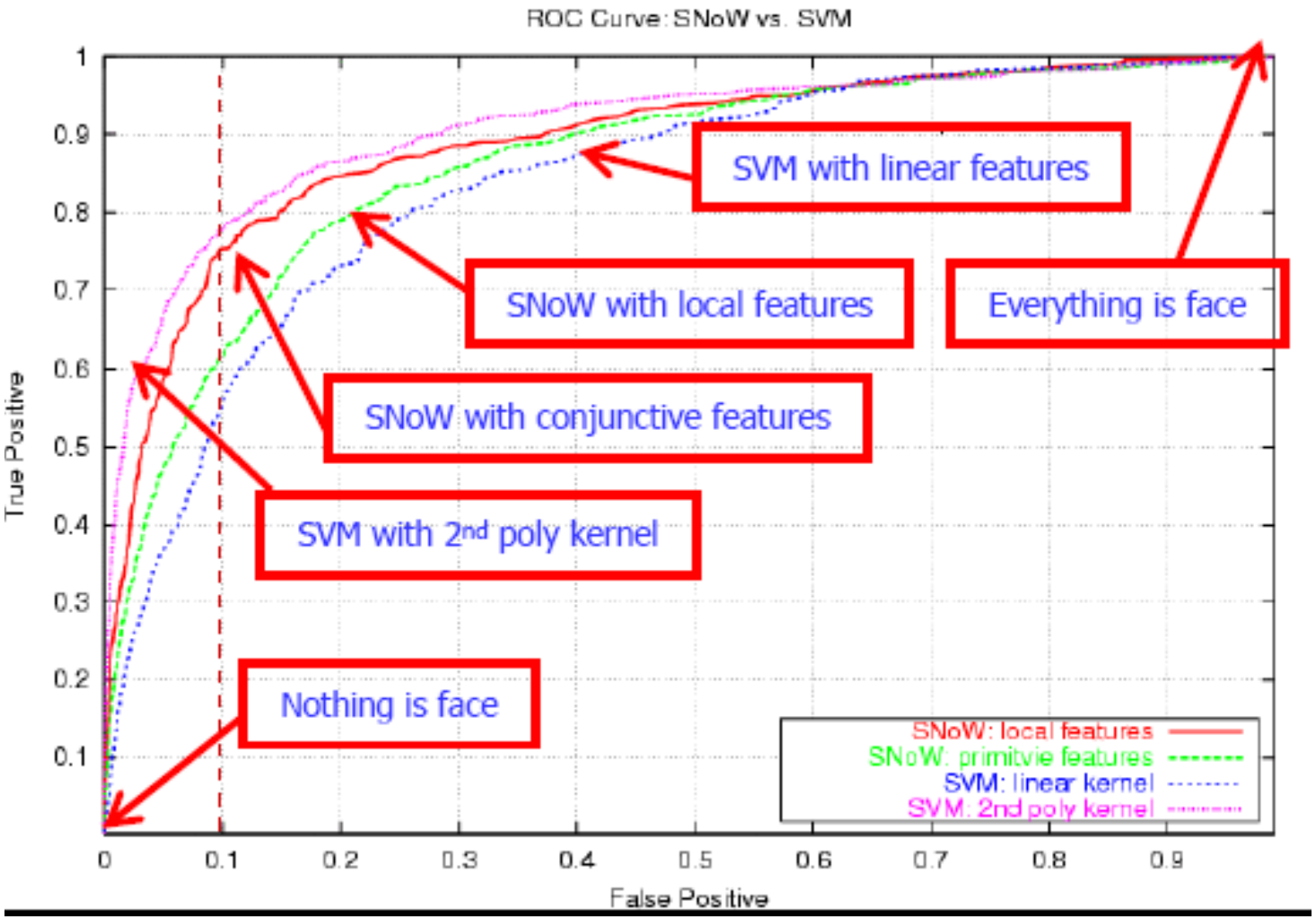
Images from: http://www.anaesthetist.com/mnm/stats/roc/

Javier Ruiz del Solar and Rodrigo Verschae, EVIC 2004

Properties of the ROC space



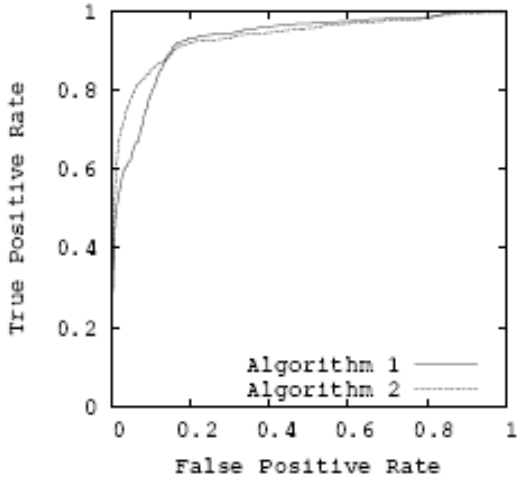
Result comparison example



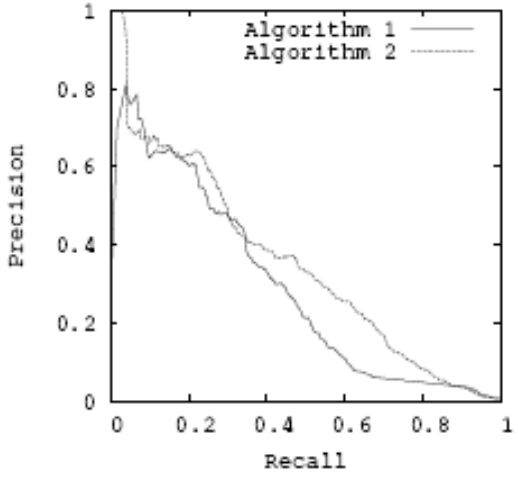
Javier Ruiz del Solar and Rodrigo Verschae, EVIC 2004

Precision-Recall Curve

•ROC curves can present an overly optimistic view of an algorithm's performance if there is a large skew in the class distribution.
•In this case Precision-Recall curves give a more informative view
•Looking at PR curves can expose differences between algorithms that are not apparent in ROC space.
•For any dataset, the ROC and PR curves for a given algorithm contain the same points.



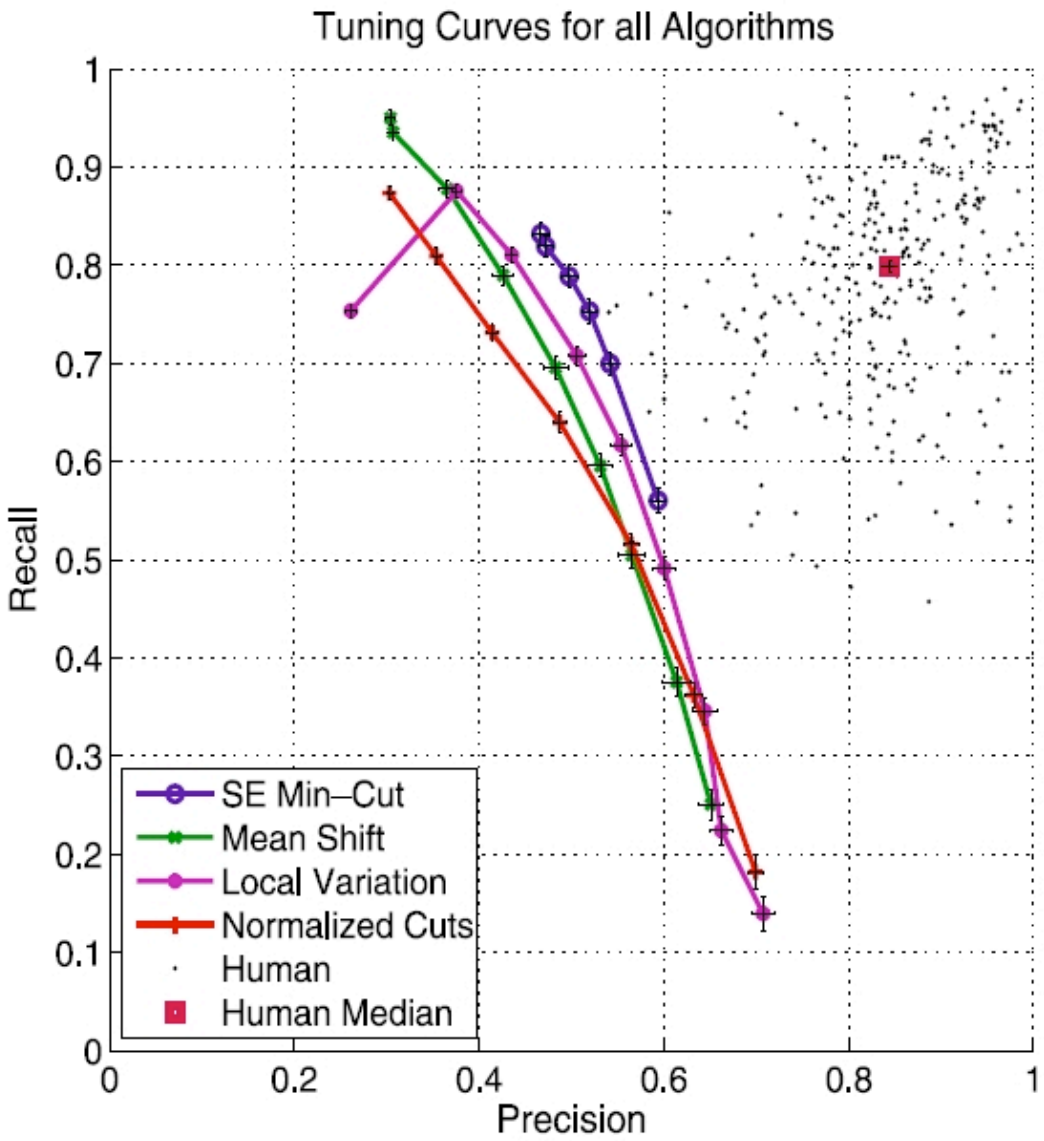
(a) Comparison in ROC space



(b) Comparison in PR space

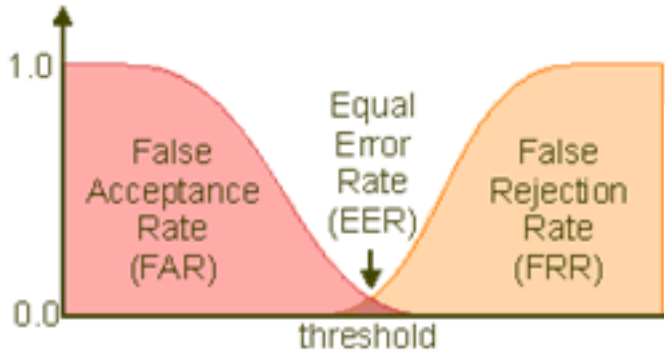
Skewness: A measure of the extent to which a probability distribution of a real-valued random variable "leans" to one side of the mean
Precision: The percentage of positive predictions that are correct.
Recall: The percentage of positive labeled instances that were predicted as positive.

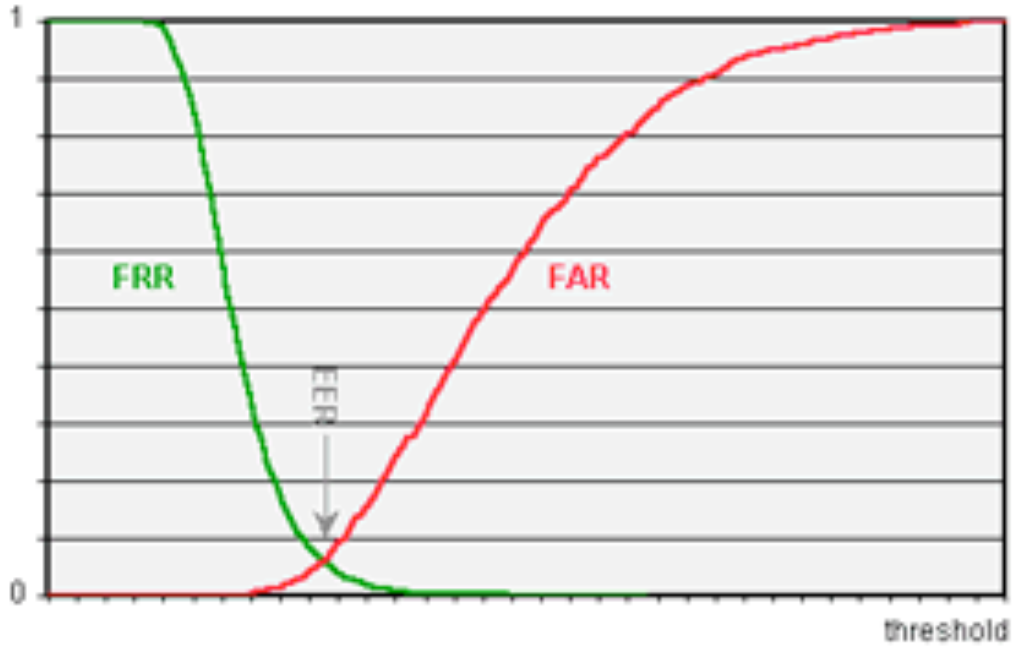
PR Curve Example



FAR-FRR Curve

Used in biometric contexts.

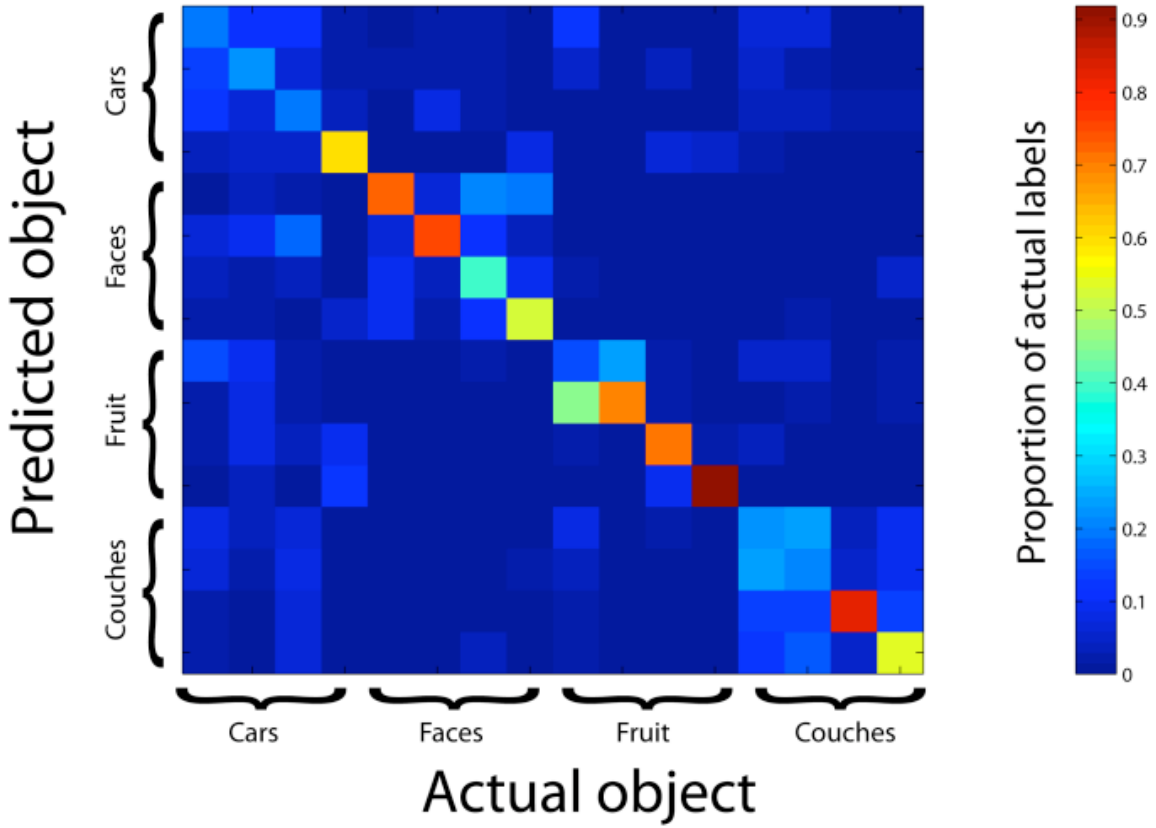




FRR=FNR
FAR=FPR

Confusion Matrix

A confusion matrix contains information about actual and predicted classifications done by a classification system.. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class.



The precision and recall values of each class can be computed individually.

Feature Selection/Extraction

1. Measurements and different from features (e.g. pixels versus area).

2. Measurement/Feature Filtering
   - Noise filtering
   - Normalization
   - Cleaning (outliers elimination, inconsistent data, duplicate data, missing values)
   - Data discretization

3. Feature/Attribute Selection
   - Selection based on attribute characteristics, e.g. variability or correlation
   - Selection using a classifier
   - Elimination versus aggregation Strategies

4. Dimensionality Reduction
   - Features/attributes can be correlated and therefore reduced using algorithms such as PCA