ROC (Receiver Operating Characteristic) curves

George C. Sakellaropoulos

ROC (Receiver Operating Characteristic) curves

- Developed in 1950s for signal detection theory to analyze noisy signals
  - Characterize the trade-off between positive hits and false alarms
- ROC curve plots the combination of sensitivity and specificity achieved when different diagnostic thresholds are applied
- In Machine Learning can be used to evaluate the performance of each classifier represented as a point on the ROC curve
  - changing various classifiers parameters changes the location of the point

ROC Analyses

Assume there are two populations, according to the presence of a disease.

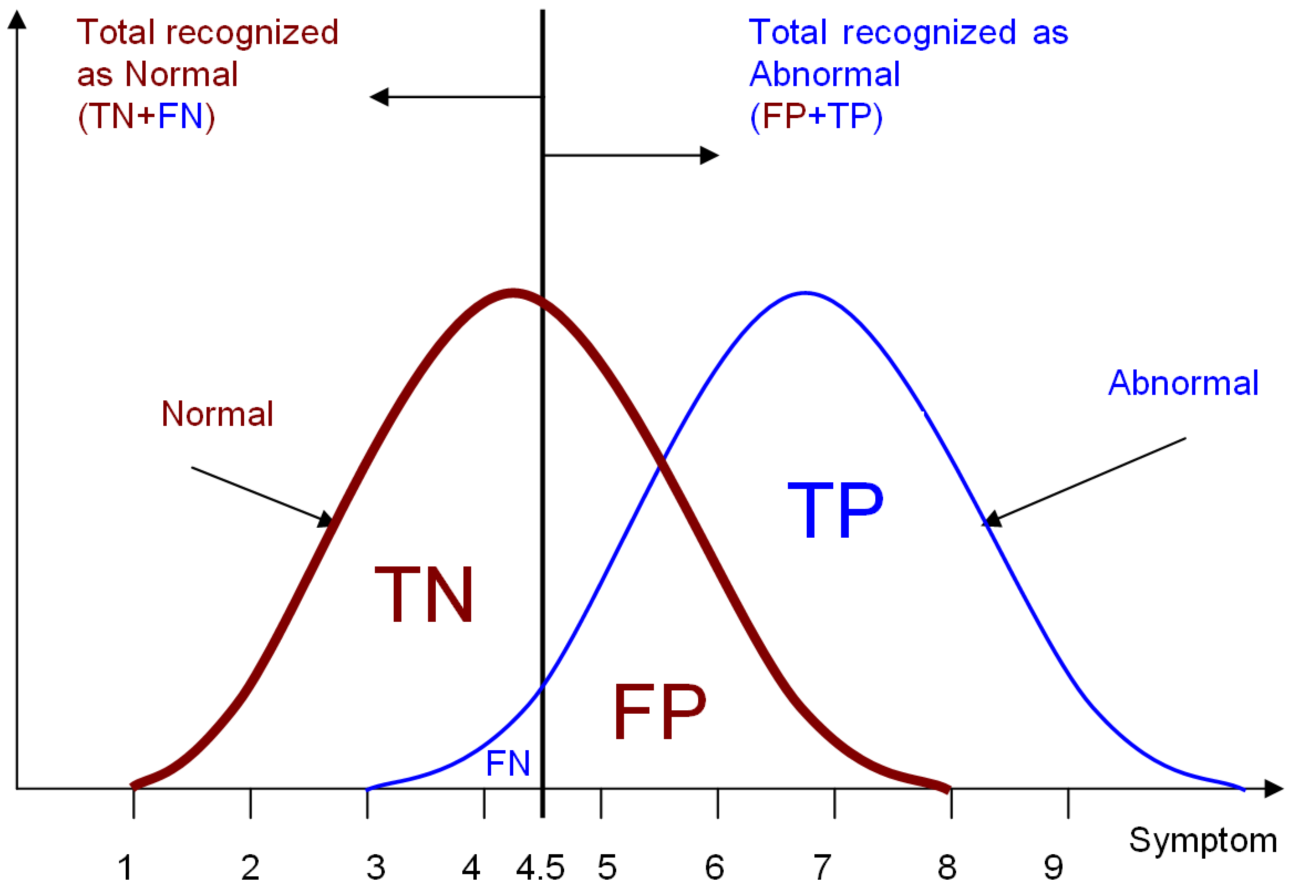
- Normal population does not have the disease
- Abnormal population has the disease

We attempt to identify the disease using the value of a diagnostic variable (also called symptom, feature, attribute). Quite often the two distributions overlap.

Specifically, we apply a threshold value, or cut-off value and we classify a person to Normal population or Abnormal population, according to his/her value of the diagnostic variable

ROC Analyses

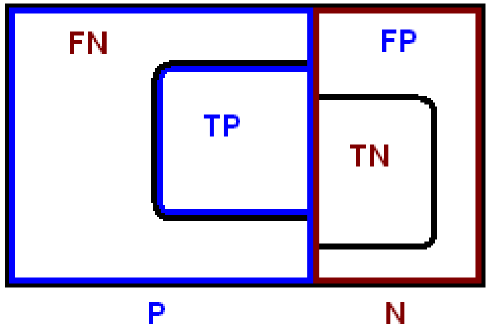
Distribution of Normal (healthy) and Abnormal (sick) patients.
Division into Normal and Abnormal populations using a threshold of 4.5

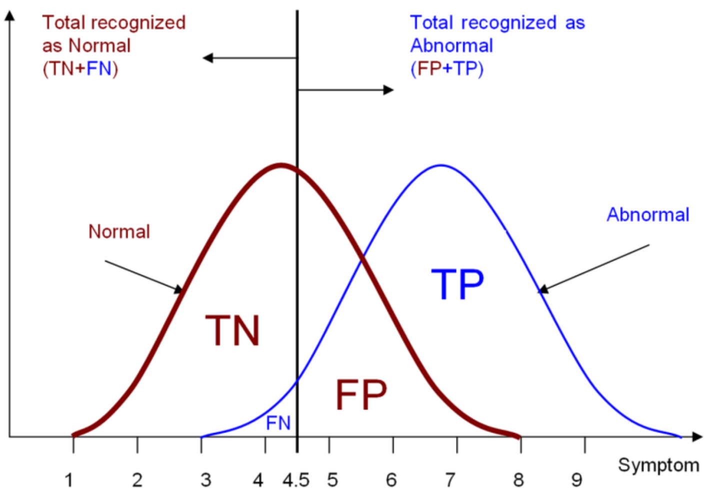


ROC Analyses

Receiver Operating Characteristics (ROC) analysis is performed by drawing curves in two-dimensional space, with axes defined by the ***TP rate*** and ***FP rate***, or equivalently, by using terms of sensitivity (=TP rate) and specificity (=1-FP rate).

- TP rate = TP/P = sensitivity
- FP rate = FP/N = 1-specificity





	Test Result		
Truth	Positive	Negative	
Positive	True Positive **TP**	False Negative **FN**	(total of true **P** positives)
Negative	False Positive **FP**	True Negative **TN**	(total of true **N** negatives)
	Total of Test recognized as Positive	Total of Test recognized as Negative	Total population

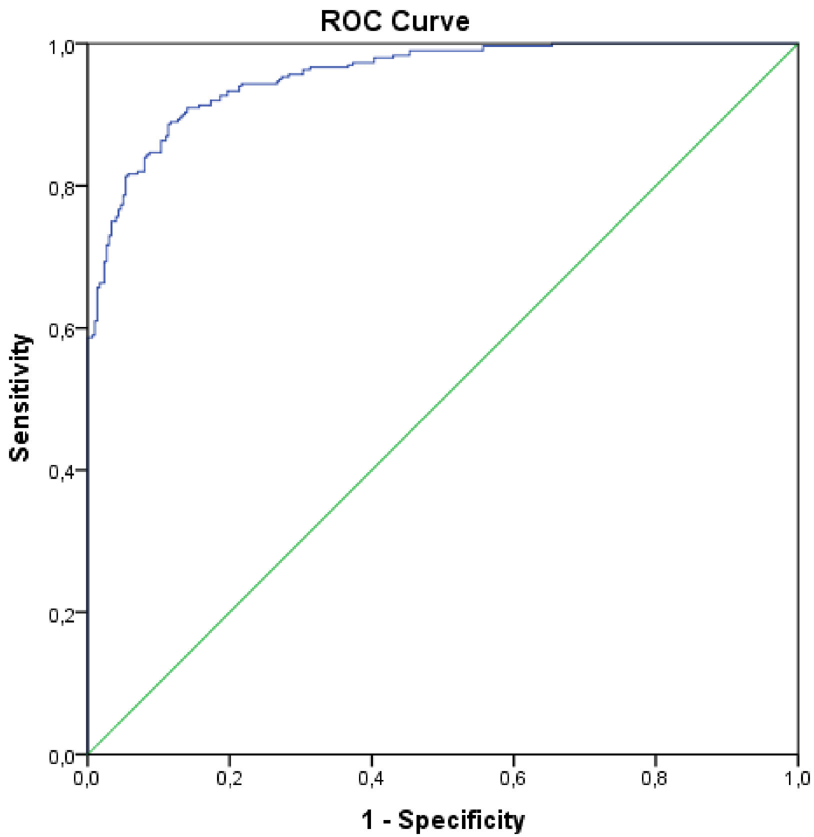
Example

- We know that a blood substance follows a distribution $N(1.32, 0.16^2)$ in the population of healthy subjects and $N(1.72, 0.18^2)$ in the population of patients with hepatitis
- We simulate 300 cases from each population, i.e. 600 cases in total. The first 300 cases are labeled as "Healthy" and the other 300 cases are labeled as "Diseased"

Group Statistics

	State.real	N	Mean	Std. Deviation	Std. Error Mean
Normal.mix	Healthy	300	1,3218	,16016	,00925
	Diseased	300	1,7288	,17743	,01024

- SPSS produces the ROC curve by evaluating [*(1-specificity) , sensitivity*] pairs for each value of the substance



Area Under the Curve

Test Result Variable(s): Normal.mix

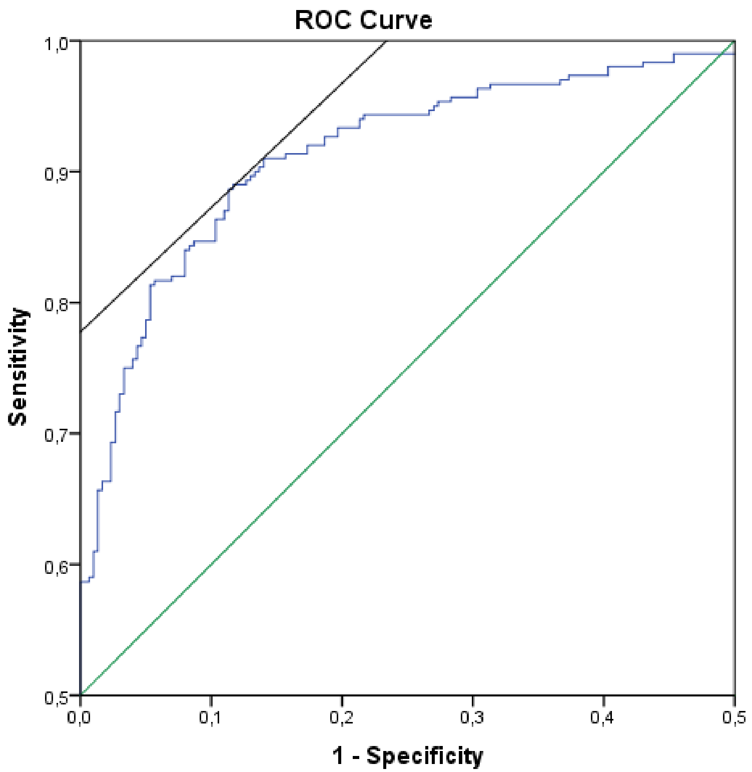
Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,955	,007	,000	,941	,969

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Finding the best cut-off value

- If we consider sensitivity and specificity as quantities of equal importance, then we can find the point as "the intercept of the ROC curve with the tangent at 45 degrees parallel to the no-discrimination line that is closest to the error-free point (0,1)"

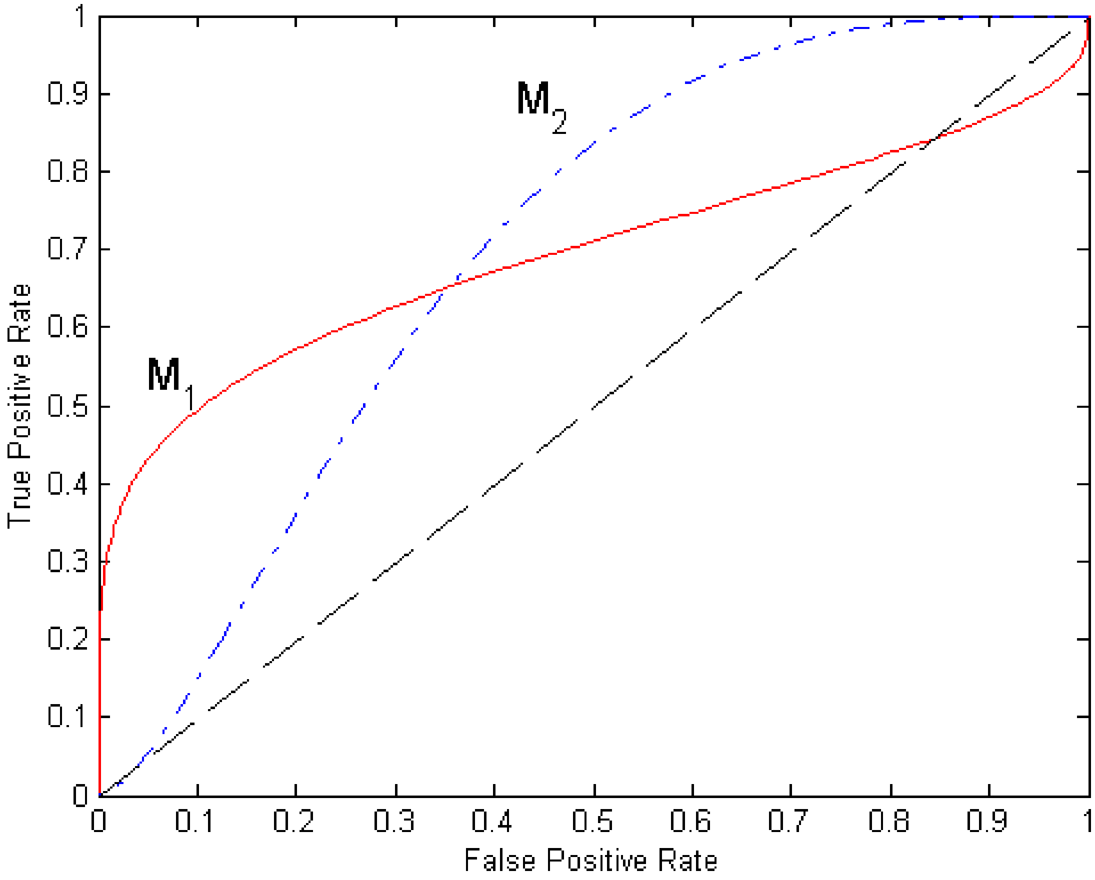


Coordinates of the Curve

Test Result Variable(s): Normal.mix

Positive if Greater Than or Equal To [a]	Sensitivity	1 - Specificity
-,1701	1,000	1,000
,8657	1,000	,997
,9194	1,000	,993
,9585	1,000	,990
1,4761	,920	,177
1,4773	,920	,173
1,4791	,917	,173
1,4807	,913	,173
1,4816	,913	,170
1,4821	,913	,167
1,4838	,913	,163
1,4856	,913	,160
1,4869	,913	,157
1,4886	,910	,157
1,4894	,910	,153
1,4897	,910	,150
1,4899	,910	,147
1,4911	,910	,143
1,4933	,910	,140
1,4952	,907	,140
1,4974	,903	,140
1,4991	,903	,137
1,5006	,900	,137
1,5028	,900	,133
1,5035	,897	,133
1,5037	,897	,130
1,5051	,893	,130
1,5065	,893	,127
1,5070	,890	,127

Using ROC for Model Comparison



How to Construct an ROC curve

Instance	P(+|A)	True Class
1	0.95	+
2	0.93	+
3	0.87	-
4	0.85	-
5	0.85	-
6	0.85	+
7	0.76	-
8	0.53	+
9	0.43	-
10	0.25	+

- Use classifier that produces posterior probability for each test instance P(+|A), where A is all available information.
- Sort the instances according to P(+|A) in decreasing order
- Apply threshold at each unique value of P(+|A)
- Count the number of TP, FP, TN, FN at each threshold
- TP rate, TPR = TP/(TP+FN)
- FP rate, FPR = FP/(FP + TN)

How to construct an ROC curve

Class	+	-	+	-	-	-	+	-	+	+	
Threshold >=	0.25	0.43	0.53	0.76	0.85	0.85	0.85	0.87	0.93	0.95	1.00
TP	5	4	4	3	3	3	3	2	2	1	0
FP	5	5	4	4	3	2	1	1	0	0	0
TN	0	0	1	1	2	3	4	4	5	5	5
FN	0	1	1	2	2	2	2	3	3	4	5
TPR	1	0.8	0.8	0.6	0.6	0.6	0.6	0.4	0.4	0.2	0
FPR	1	1	0.8	0.8	0.6	0.4	0.2	0.2	0	0	0

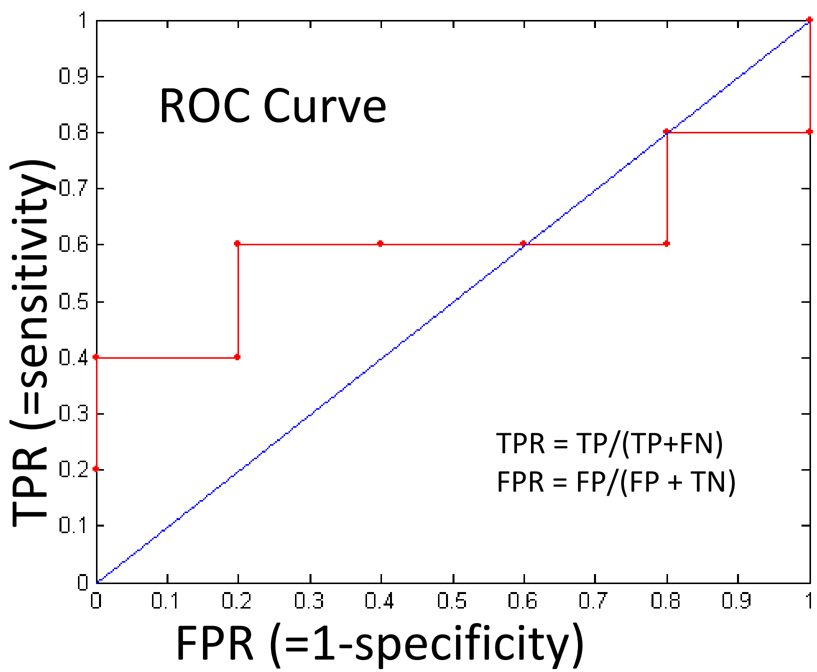
Instance	P(+|A)	True Class
1	0.95	+
2	0.93	+
3	0.87	-
4	0.85	-
5	0.85	-
6	0.85	+
7	0.76	-
8	0.53	+
9	0.43	-
10	0.25	+

TP

FP

FN

TN



ROC Analyses

The ideal model/classifier would be one represented by a location (0,1) on the graph, corresponding to 100% specificity and 100% sensitivity

Points (0,0) and (1,1) represent 100% specificity and 0% sensitivity for the first point, and 0% specificity and 100% sensitivity for the second, respectively.

Neither of the two points would represent an acceptable model. All points lying on the curve connecting the two points ((0,0) and (1,1)) represent random guessing of the classes (equal values of 1-specificity and sensitivity, or in other words, equal values of TP and FP rates.

That means that these models/classifiers would recognize equal amounts of TP and FP, with the point at (0.5, 0.5) representing 50% specificity and 50% sensitivity.

These observations suggests the strategy for always “operating” in the region above the diagonal line (y=x), since in this region the TP rate is higher than the FP rate.

ROC Analyses

How to decide which of the two classifiers constitutes a better model/classifier of the data?

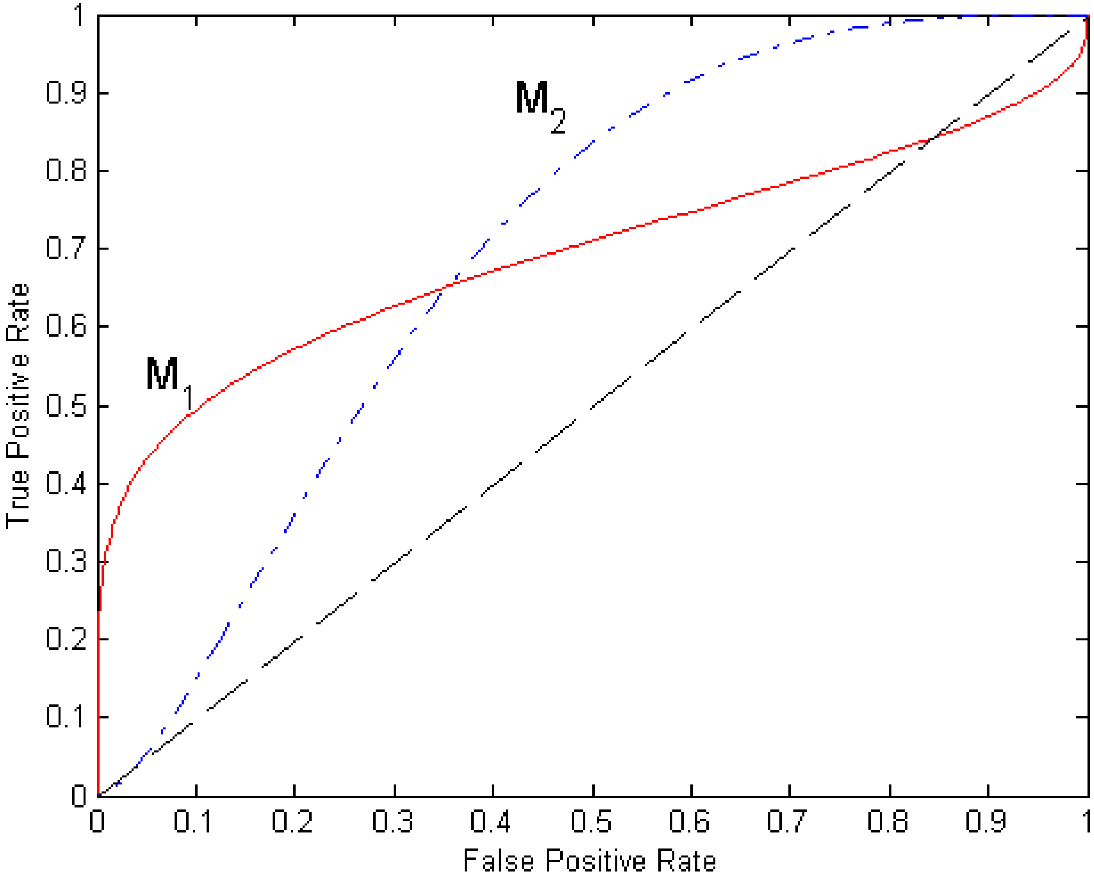
We could perform visual analysis: the curve more to the upper left would indicate a better classifier. However, the curves often overlap, and a decision may not be so easy to make.

A popular method used to solve the problem is called Area Under Curve (AUC).

The area under the diagonal curve is 0.5. Thus, we are interested in choosing a model/classifier which has maximum area under its corresponding ROC curve: the larger the area the better performing the model/classifier is.

There exist other measures similar to the AUC for assessing the goodness of a model/classifier

Using ROC for Model Comparison



- ROC plots allow for visual comparison of several models (classifiers).
- No model consistently outperform the other
  - M_1 is better for small FPR
  - M_2 is better for large FPR
- Area Under the ROC curve
  - Ideal:
    - Area = 1
  - Random guess:
    - Area = 0.5

END