Model evaluation

UNIVERSIDAD
DE ANTIOQUIA
1803

Nestór Rendón Gabriel Maldonado

Statistical Learning

Model Evaluation

Model evaluation metrics are required to quantify model performance.

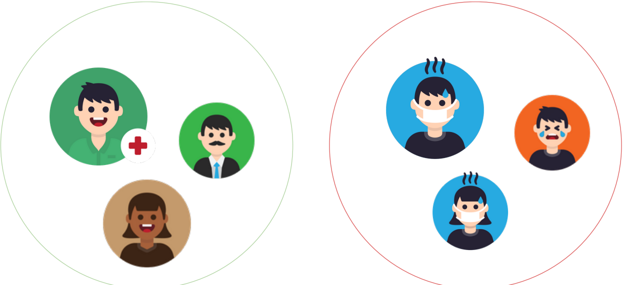
The Accuracy is the proportion of the total number of predictions that were correct.

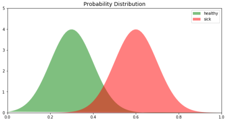
Let be a fitted model f(x), and the aim is to see how well it performs. The prediction accuracy for the model is very very high, at 99.9 % . But what if the costs of having a mis-classified actual positive (or false negative) is very high.

The accuracy is not the be-all and end-all model metric to use when selecting the best model
So there are different methods to compare the performance of models..

Model Evaluation

$D = (x_i, y_i)$ where i=1,2,...,n, and $y = (1 - 0)$



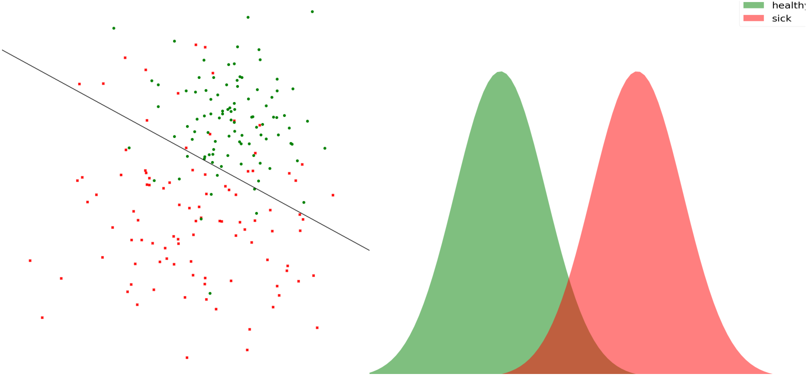


Confusion Matrix

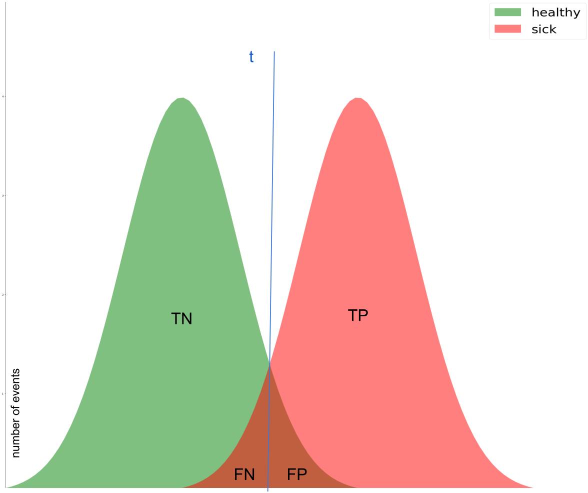
The root of a dichotomous decision process is a threshold (t)-based rule on a continuous variable, y, that will drive the decision, D, as positive or negative according to

$$D = \begin{cases} +, & \text{if } y >= t. \\ -, & \text{otherwise.} \end{cases} \tag{1}$$

Confusion Matrix



Confusion Matrix



Confusion Matrix

The Clasification rate (Accuracy) is the proportion of correct decisions to total decisions – as a measure of the goodness of the rule

$$CR = \frac{(TN + TP)}{n} \tag{2}$$

The Missclasification rate

$$MCR = \frac{(FN + FP)}{n} \tag{3}$$

Where n is the number of data.

$$n = TN + TP + FN + FP \tag{4}$$

Confusion Matrix

Karl Pearson. He used the term Contingency Table (1904).
During War World 2, Detection Theory was developed as investigation of the relations between stimulus and responds. The confusion matrix was used there.
The term was used in psychology.

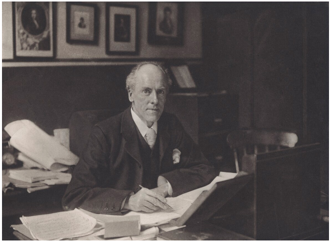


Figure: Karl Pearson

Confusion Matrix

	Predicted		
Observed		FALSE	TRUE
	FALSE	True Negative	False Positive
	TRUE	False Negative	True Positive

Confusion Matrix

Metrics:
When the Class is positive, which are correctly classified?
The correct positive fraction, Sensitivity, true positive rate, hit rate, recall:

$$CPF = \frac{TP}{TP + FN} \tag{5}$$

Recall shall be the model metric we use to select our best model when there is a high cost associated with False Negative.

Confusion Matrix

The False positive rate, :

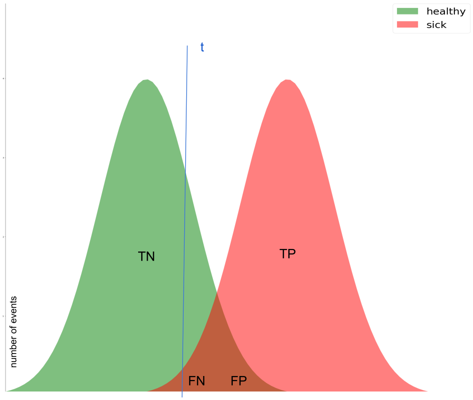
$$FPR = \frac{FP}{FP + TN} \tag{6}$$

When the Class is negative, whiTrue Negativech are correctly classified?
The correct negative fraction, Specificity,True negative rate:

$$CNF = \frac{TN}{TN + FP} = 1 - FPR \tag{7}$$

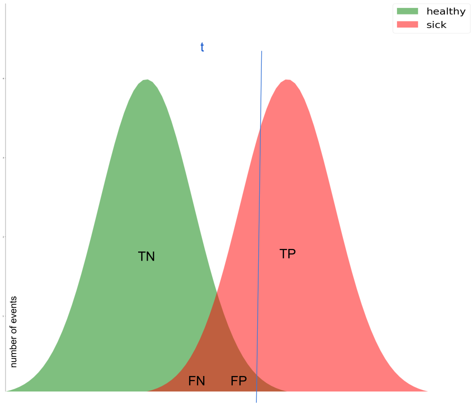
Confusion Matrix

Sencibility increases, Specifity decreases (Virus,Cancer)



Confusion Matrix

Sencibility decreases, Specifity increases (SPAM)



ROC

Receiver Operating Characteristics (ROC)

"The receiver operating characteristic (ROC) curve was introduced in World War II military radar operations as a means to characterize the operators' ability to correctly identify friendly or hostile aircraft based on a radar signal." Brown, Herbert (2005)

"ROC graphs are two-dimensional graphs in which tp rate is plotted on the Y axis and fp rate is plotted on the X axis. An ROC graph depicts relative tradeoffs between benefits (true positives) and costs (false positives)." Fawcett (2005)

ROC

If the threshold decreases the sensitivity increases the specificity increases

If the threshold increases the sensitivity decreases the specificity increases

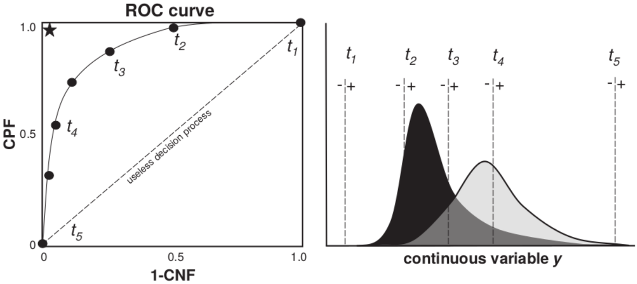
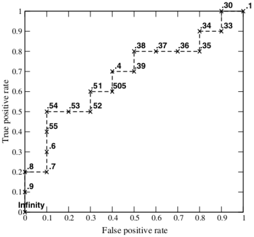
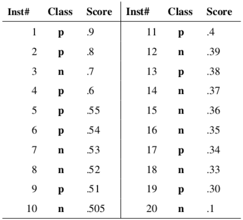


Figure: Roc Curve,C.D. Brown, H.T. Davis (2006)

ROC





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

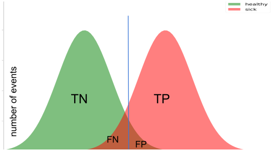
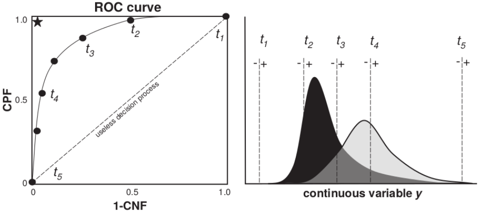


Figure: T. Fawcett / Pattern Recognition Letters 27 (2006)

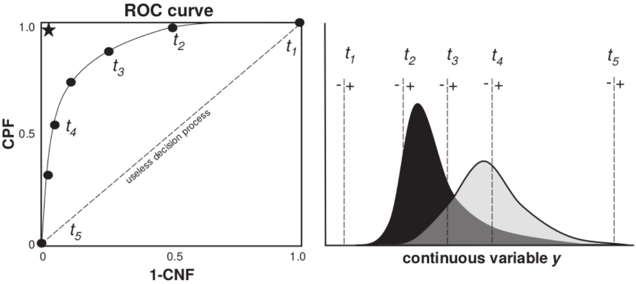
ROC

1. $\star$, Best case in which $CPF = 1$ and $CNF = 1$
2. t1, unconditionally issuing positive classifications
3. t2 Positive classification only weak evidence.
4. t4 Positive clasification only strong evidence
5. t5, classifier commits no false positive errors but also gains no true positives.

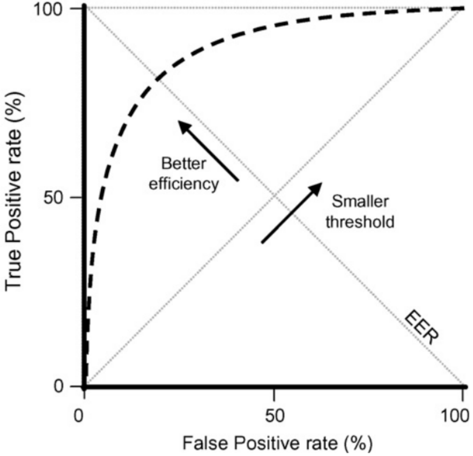


ROC

1. Any classifier that appears in the lower right triangle performs worse than random guessing
2. Any classifier on the diagonal may be said to have no information about the class



ROC


100
50
0
0
50
100
True Positive rate (%)
False Positive rate (%)
Better efficiency
Smaller threshold
EER

ROC

Advantages

The ROC is also invariant monotonic transformations (if x>y, then f(x)>f(y)), which makes it a convenient representation insensitive to scale.

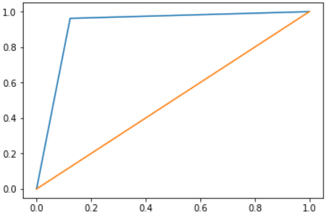
In applications where detecting positive events is critical the analyst can very rapidly examine the ROC curve for a decision process and assess the minimum incorrect positive.

ROC

Note: The literature reflects many parametric, semiparametric, and nonparametric estimation methods have been proposed for estimating the ROC curve and its associated summary measures. L. Gon¸calves et al (2014)

ROC

For a model which gives class as output, will be represented as a single point in ROC plot



AUC

The area under the curve (AUC)is considered as summarie of the discriminatory accuracy of a test is given by:

$$AUC = \int_0^1 ROC(u)du \tag{8}$$

.90-1 = excellent (A)

.80-.90 = good (B)

.70-.80 = fair (C)

.60-.70 = poor (D)

.50-.60 = fail (F)

F1-score

Of a particular class how many were correctly predicted? Precision is a good measure to determine, when the costs of False Positive is high.

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

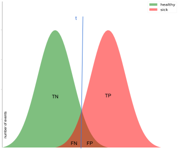
F1-score

Remember Recall or Sensitivity is the correct positive fraction,

$$CPF = \frac{TP}{TP + FN} \tag{10}$$

Of a particular class how many were correctly predicted?

$$Precision = \frac{TP}{TP + FP} \tag{11}$$



F1-score

The F1 score is calculated based on the precision and recall of each class.

$$F1score = 2 * \frac{Precision * Recall}{Precision + recall} \tag{12}$$

F1 is usually more useful than accuracy, especially if you have an uneven class distribution.

Thanks