Confusion Matrix

A confusion matrix (Kohavi and Provost, 1998) contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. The following table shows the confusion matrix for a two class classifier.

The entries in the confusion matrix have the following meaning in the context of our study:

- *a* is the number of **correct** predictions that an instance is **negative**,
- *b* is the number of **incorrect** predictions that an instance is **positive**,
- *c* is the number of **incorrect** of predictions that an instance **negative**, and
- *d* is the number of **correct** predictions that an instance is **positive**.

		Predicted	
		Negative	Positive
Actual	Negative	**a**	**b**
	Positive	**c**	**d**

Several standard terms have been defined for the 2 class matrix:

- The *accuracy* (*AC*) is the proportion of the total number of predictions that were correct. It is determined using the equation:

$$AC = \frac{a+d}{a+b+c+d}$$
[1]

- The *recall* or *true positive rate* (*TP*) is the proportion of positive cases that were correctly identified, as calculated using the equation:

$$TP = \frac{d}{c+d}$$
[2]

- The *false positive rate* (*FP*) is the proportion of negatives cases that were incorrectly classified as positive, as calculated usingthe equation:

$$FP = \frac{b}{a+b}$$
[3]

- The *true negative rate* (*TN*) is defined as the proportion of negatives cases that were classified correctly, as calculated using the equation:

$$TN = \frac{a}{a+b}$$
[4]

- The *false negative rate* (*FN*) is the proportion of positives cases that were incorrectly classified as negative, as calculated using the equation:

$$FN = \frac{c}{c+d}$$
[5]

- Finally, *precision* (*P*) is the proportion of the predicted positive cases that were correct, as calculated using the equation:

$$P = \frac{d}{b+d}$$
[6]

The accuracy determined using equation 1 may not be an adequate performance measure when the number of negative cases is much greater than the number of positive cases (Kubat et al., 1998). Suppose there are 1000 cases, 995 of which are negative cases and 5 of which are positive cases. If the system classifies them all as negative, the accuracy would be 99.5%, even though the classifier missed all positive cases. Other performance measures account for this by including *TP* in a product: for example, *geometric mean* (*g-mean*) (Kubat et al., 1998), as defined in equations 7 and 8, and *F-Measure* (Lewis and Gale, 1994), as defined in equation 9.

$$g-mean_1 = \sqrt{TP * P}$$
[7]

$$g-mean_2 = \sqrt{TP * TN}$$
[8]

$$F = \frac{(\beta^2 + 1) * P * TP}{\beta^2 * P + TP}$$
[9]

In equation 9, β has a value from 0 to infinity and is used to control the weight assigned to *TP* and *P*. Any classifier evaluated using equations 7, 8 or 9 will have a measure value of 0, if all positive cases are classified incorrectly.

Another way to examine the performance of classifiers is to use a ROC graph, described on the next page.