



Why

We want to summarize the (label, prediction) pairs for a particular classifier on a particular dataset in a two-dimensional table, or matrix.

Definition

Let A be a nonempty set and $B = \{-1, 1\}$. For a dataset $(a^1, b^1), \dots, (a^n, b^n)$ in $A \times B$, and classifier $G : A \rightarrow B$, the *confusion matrix* C is defined

$$C = \begin{bmatrix} \# \text{ true negatives} & \# \text{ false negatives} \\ \# \text{ false positives} & \# \text{ true positives} \end{bmatrix} = \begin{bmatrix} C_{\text{tn}} & C_{\text{fn}} \\ C_{\text{fp}} & C_{\text{tp}} \end{bmatrix}.$$

Using this notation, $C_{\text{tn}} + C_{\text{fn}} + C_{\text{fp}} + C_{\text{tp}} = n$. $N_{\text{n}} := C_{\text{tn}} + C_{\text{fp}}$ is the number of negative examples. $N_{\text{p}} := C_{\text{fn}} + C_{\text{tp}}$ is the number of positive examples.

The diagonal elements of the confusion matrix give the numbers of correct predictions. The off-diagonal entries give the numbers of incorrect predictions for the two types of errors (see **Classifier Errors**).

In this notation, the false positive rate is C_{fp}/n , the false negative rate is C_{fn}/n and the error rate is the sum of these, $(C_{\text{fn}} + C_{\text{fp}})/n$.

The true positive rate is $C_{\text{tp}}/(C_{\text{fn}} + C_{\text{tp}})$. The true negative rate is $C_{\text{tn}}/(C_{\text{tn}} + C_{\text{fp}})$. The false alarm rate is $C_{\text{fp}}/(C_{\text{tn}} + C_{\text{fp}})$. The precision is $C_{\text{tp}}/C_{\text{tp}} + C_{\text{fp}}$

