**Course: Advanced Bioinformatics**

**Module title: Confusion Matrix**

**Module no. : 175**

In the field of machine learning, a confusion matrix, also known as a contingency table or an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one (in unsupervised learning it is usually called a matching matrix). Each column of the matrix represents the instances in a predicted class while each row represents the instances in an actual class (or vice-versa). The name stems from the fact that it makes it easy to see if the system is confusing two classes (i.e. commonly mislabeling one as another).

Table of confusion

In predictive analytics, a table of confusion (sometimes also called a confusion matrix), is a table with two rows and two columns that reports the number of false positives, false negatives, true positives, and true negatives. This allows more detailed analysis than mere proportion of correct guesses (accuracy). Accuracy is not a reliable metric for the real performance of a classifier, because it will yield misleading results if the data set is unbalanced (that is, when the number of samples in different classes vary greatly). For example, if there were 95 cats and only 5 dogs in the data set, the classifier could easily be biased into classifying all the samples as cats. The overall accuracy would be 95%, but in practice the classifier would have a 100% recognition rate for the cat class but a 0% recognition rate for the dog class.

Assuming the confusion matrix above, its corresponding table of confusion, for the cat class, would be:

|  |  |
| --- | --- |
| 5 true positives (actual cats that were correctly classified as cats) | 2 false positives (dogs that were incorrectly labeled as cats) |
| 3 false negatives (cats that were incorrectly marked as dogs) | 17 true negatives (all the remaining animals, correctly classified as non-cats) |

The final table of confusion would contain the average values for all classes combined.

Let us define an experiment from P positive instances and N negative instances for some condition. The four outcomes can be formulated in a 2×2 contingency table or confusion matrix, as follows:

