**Multiclass Confusion Matrix in Python: Clarity without Confusion.**

In the vast field of Machine Learning, the general focus is to predict an outcome using the available data. The prediction task is also called "classification problem" when the outcome represents different classes, otherwise is called "regression problem" when the outcome is a numeric measurement.1 As regards to classification, the most common setting involves only two classes, although there may be more than two.1, 2

Classification tasks in machine learning involving more than two classes are known by the name of "multi-class classification". Performance indicators are very useful when the aim is to evaluate and compare different classification models or machine learning techniques. Many metrics come in handy to test the ability of a multi-class classifier.1, 3

A confusion matrix is a tabular way of visualizing the performance of your prediction model. Each entry in a confusion matrix denotes the number of predictions made by the model where it classified the classes correctly or incorrectly.2, 3

The confusion matrix is a cross table that records the number of occurrences between two raters, the true/actual classification and the predicted classification. For consistency reasons throughout the paper, the columns stand for model prediction whereas the rows display the true classification.1

The classes are listed in the same order in the rows as in the columns, therefore the correctly classified elements are located on the main diagonal from top left to bottom right and they correspond to the number of times the two raters agree.1

In other words: A confusion matrix, as the name suggests, is a matrix of numbers that tell us where a model gets confused. It is a class-wise distribution of the predictive performance of a classification model - that is, the confusion matrix is an organized way of mapping the predictions to the original classes to which the data belong. This also implies that confusion matrices can only be used when the output distribution is known.4

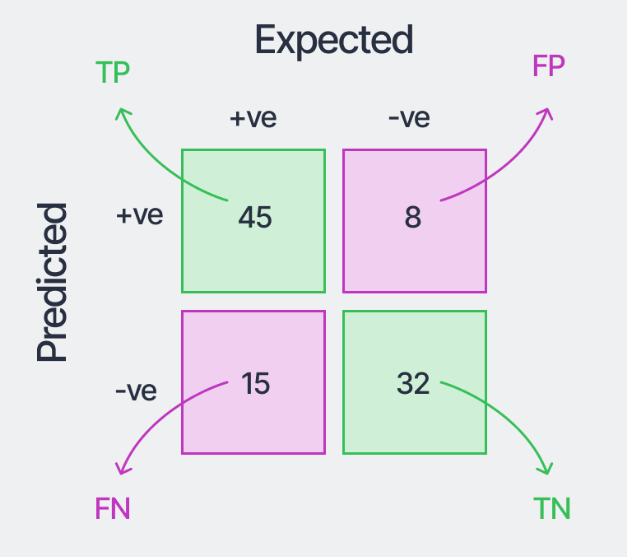
Firstly, we are going to understand the confusion matrix for binary classification problems and the calculation of its metrics, and then give some clarity to the issue of multiclass confusion matrix, the calculation of its metrics and the ways to calculate its macro and micro averages.

**Confusion Matrix for Binary Classes:**

A binary class dataset is one that consists of just two distinct categories of data. These two categories can be named the “positive” and “negative” for the sake of simplicity. Suppose we have a binary class imbalanced dataset consisting of 60 samples in the positive class and 40 samples in the negative class of the test set, which we use to evaluate a machine learning model.

Now, to fully understand the confusion matrix for this binary class classification problem, we first need to get familiar with the following terms:2, 3, 4

* True Positive (TP) refers to a sample belonging to the positive class being classified correctly.
* True Negative (TN) refers to a sample belonging to the negative class being classified correctly.
* False Positive (FP) refers to a sample belonging to the negative class but being classified wrongly as belonging to the positive class.
* False Negative (FN) refers to a sample belonging to the positive class but being classified wrongly as belonging to the negative class.4

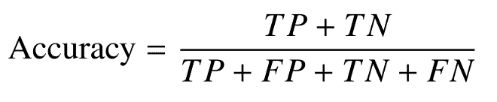


Confusion Matrix for a binary class dataset.4

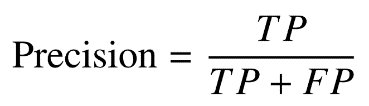
An example of the confusion matrix we may obtain with the trained model is shown above for this example dataset. This gives us a lot more information than just the accuracy of the model. Adding the numbers in the first column, we see that the total samples in the positive class are 45+15=60. Similarly, adding the numbers in the second column gives us the number of samples in the negative class, which is 40 in this case. The sum of the numbers in all the boxes gives the total number of samples evaluated. Further, the correct classifications are the diagonal elements of the matrix - 45 for the positive class and 32 for the negative class. Now, 15 samples (bottom-left box) that were expected to be of the positive class were classified as the negative class by the model. So it is called “False Negatives” because the model predicted “negative,” which was wrong. Similarly, 8 samples (top-right box) were expected to be of negative class but were classified as “positive” by the model. They are thus called “False Positives.” We can evaluate the model more closely using these four different numbers from the matrix.4

In general, we can get the following quantitative evaluation metrics from this binary class confusion matrix:4

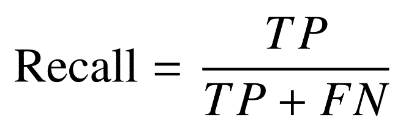
1. Accuracy: The number of samples correctly classified out of all the samples present in the test set.



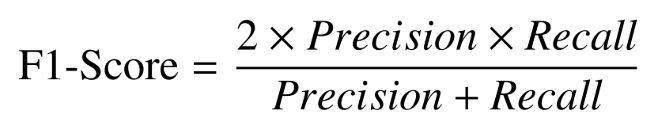
1. Precision (for the positive class): The number of samples actually belonging to the positive class out of all the samples that were predicted to be of the positive class by the model.



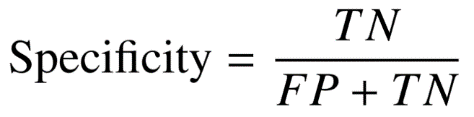
1. Recall (for the positive class): The number of samples predicted correctly to be belonging to the positive class out of all the samples that actually belong to the positive class.



1. F1-Score (for the positive class): The harmonic mean of the precision and recall scores obtained for the positive class.



1. Specificity: The number of samples predicted correctly to be in the negative class out of all the samples in the dataset that actually belong to the negative class.

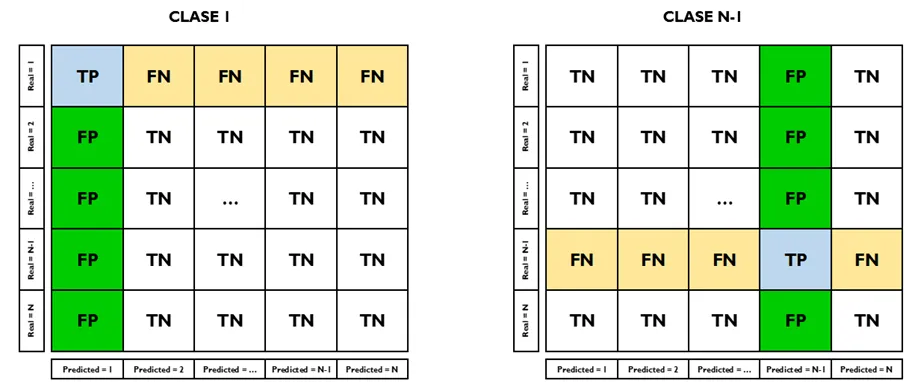


Now that you have better understood the confusion matrix for binary classification models, let's finally give some clarity and understand what the confusion matrix for multi-class models is all about.

**Confusion Matrix for Multiple Classes:**

In some classification problems we may encounter a model with a number of labels to be classified greater than 2, and the precision formula cannot be taken from the previous expression. When we have a classification model involving N labels (where N>2) the confusion matrix (MC) has NxN dimensions.5

The concept of the multi-class confusion matrix is similar to the binary-class matrix. The columns represent the original or expected class distribution, and the rows represent the predicted or output distribution by the classifier.4



Confusion Matrix for a Multiclass Classifier.5

In this figure, each column represents the False Positives (FP) and the row represents the False Negatives (FN) for that label. Note that where the index of the row converges with the column corresponds to the True Positive (TP), that is, where the estimate of our model coincides with reality, and the rest constitute the True Negatives (TN).5

Having clarified this, it only remains for us to clarify that the metrics are calculated following the same formulas, changing only in terms of the estimation of the TP, TN, FP, FN and in terms of the way of calculating the average of the metrics.

There are different ways to calculate accuracy, precision, and recall for multi-class classification. You can calculate metrics by each class or use macro - or micro - averaging. A suitable metric depends on the specific problem and the importance of each class or instance.6

Finally, I share with you a very basic Python code to graph a multiclass confusion matrix and calculate its metrics. (github.com)

References:

1. Grandini M, Bagli E. Metrics for Multi-Class Classification: an Overview. arXiv Vanity [Internet]. 2023-02-14. [Citado: 2023-12-04]:24. Disponible en: <https://www.arxiv-vanity.com/papers/2008.05756/>.

2. Mohajon J. Confusion Matrix for Your Multi-Class Machine Learning Model. Newsletter Towards Data Science [Internet]. 2023. [Citado: 2023-11-04]:13. Disponible en: <https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826>.

3. Publishing Group. How to interpret a confusion matrix for a machine learning model. Newsletter Evidently AI, [Internet]. 2023. [Citado: 2023-11-04]:17. Disponible en: <https://www.evidentlyai.com/classification-metrics/confusion-matrix>.

4. Kundu R. Confusion Matrix: How To Use It & Interpret Results [Examples]. Newsletter V7labs [Internet]. 2022-09-13. [Citado: 2023-12-04]:25. Disponible en: <https://www.v7labs.com/blog/confusion-matrix-guide>.

5. Barrios Bustamante W. Calculando la precisión en un modelo de Clasificación Multiclase. Medium.com [Internet]. 2021-04-02. [Citado: 2023-12-04]:12. Disponible en: <https://wbarriosb.medium.com/calculando-la-precisi%C3%B3n-en-un-modelo-de-clasificaci%C3%B3n-multiclase-224d96f52043>.

6. Publishing Group. Accuracy, precision, and recall in multi-class classification. Newsletter Evidently AI, [Internet]. 2023. [Citado: 2023-11-04]:25. Disponible en: <https://www.evidentlyai.com/classification-metrics/multi-class-metrics>.