Accuracy

In general, accuracy is the number of items well classified over the total number of inputs. If we want to evaluate the accuracy over a class, we need to evaluate the number of items of such class well classified over the total of inputs of such class.

Precision

The precision is evaluated over the class and it is the number of items that were well classified of such class over the total of predictions that denoted this class.

Recall

The Recall is evaluated over the class and it is the number of items that were well classified of such class over the real number of inputs of this class.

F1-Score

Normally a model in the way how it gets more precision it starts reducing the recall the ideal is to have a balance. F1-Score follows the following formula:

F1-Score = 2 \* (precision \* recall) / (precision + recall)

The highest value will be F1-Score = precision = recall. In other words, F1-Score measure the valance between precision and recalls of a class prediction.

Macro-F1

Macro-F1 is the average value of all the F1-Score values of all the classes to predict.

Hamming Loss

Hamming Loss is the number of items wrong predicted over the total of inputs.

Jaccard Score

Jaccard Score or Jaccard similarity coefficient score is the relationship for a class between the times that it was well predicted against all the times that class appears in the classifications. For example if we have one input like {0,1,2,2} and the prediction is the following {0,2,1,2} for the class number 2 the correct predictions is just one, and the times that appear in the classification as input or as prediction is 3 then the jaccard score for the class 2 is 1/3. We can get the average Jaccard score for all the classes and get a value for the entire model called Jaccard Score macro. Understanding this we can understand Jaccard score as the similarity between the inputs and the predictions considering 0% the minimum when inputs and predictions are totally different and 100% when the accuracy is 100%.

Log loss

Log loss or cross-entropy is a metric quite useful when the prediction is based in probabilities. In general, log loss is the error between the input and the output. Even if the accuracy is 100% if we use prediction based on probabilities there will always be a gap between the input and the output for example if have one image that is dog and a classifier that differs between cat and dogs and it says that the image is 90% dog and 10% cat, then it will be well classified but the error is of 10%. It is applied a log function to avoid handling too small numbers.