Evaluation Metrics Guide

Evaluation Metrics Definitions

IoU grouping parameter is used to discard linkage of wrong label between ground truth and prediction having IoU less than IoU Grouping

Generally, positive/negative defines whether we have a prediction or not and True/False defines whether our prediction is True/False

	Prediction	**Reality**
True Positive (TP)	Positive	Positive
True Negative (TN)	Negative	Negative
False Positive (FP)	Positive	Negative
False Negative (FN)	Negative	Positive

Accuracy: $\frac{Correct\ answers}{All\ Posibilities} = \frac{TP+TN}{TP+TN+FP+FN}$

- Measure of how correct our Data is.
- Best = 1 ----- Worst = 0

Precision: $\frac{Predicted\ Correctly}{All\ Predicted} = \frac{TP}{TP+FP}$

- Measure the fraction of relevant examples among all of the examples predicted that belong to a certain class.
- Best = 1 ----- Worst = 0

Recall: $\frac{Predicted\ Correctly}{All\ that\ should\ be\ Predicted} = \frac{TP}{TP+FN}$

- Measure the fraction of relevant examples which were predicted to belong to a class with respect to all examples that truly belong to a class.
- Best = 1 ----- Worst = 0

F-Score: $\frac{(1+\beta^2) \times Precision \times Recall}{(\beta^2 \times Precision)+Recall}$

- Evaluate machine learning model in one number by combining precision and recall
- A good f-score means that the model is effective in terms of how precisely it classifies the data and that it covers a good proportion of the cases that it should have classified correctly.
- Best = 1 ----- Worst = 0
    - If ↑ = 1 ✉ neutral
    - If ↑ < 1 ✉ focuses more on precision
    - If ↑ > 1 ✉ focuses more on recall

- Before Continuing, we need to give an example:

- Let us take a case where we have 95% of the population does not have cancer and 5% of the population have cancer.
- If over 100 person we detect 100 persons that do not have cancer, then we will have a case where accuracy = 95%.
- But in reality, we need this model to predict if a person has cancer or not so these results are unacceptable, because we cannot find any of the persons that has cancer.
- Another example is the case where 95% of data should be detected but we detect 100% of the data True. In this case we will have precision = 95% and Recall = 100% but the purpose of finding the 5% makes the model be very bad.
- This is why a new metric will be introduced (balanced accuracy), in which we are taking into consideration how the dataset is distributed or balanced

---

True Negative Rate: $\frac{TN}{TN+FP}$

True Positive Rate: $\frac{TP}{TP+FN}$

Balanced Accuracy: $\frac{TNR+TPR}{2}$

- This parameter takes into consideration how balanced is the data.