**Evaluating Model Performance**

**Metrics**

We can use **F-beta score** as a metric that considers both precision and recall:

Fβ=(1+β2)⋅precision⋅recall(β2⋅precision)+recall

In particular, when β=0.5, more emphasis is placed on precision. This is called the **F0.5 score** (or F-score for simplicity).

**Note: Recap of accuracy, precision, recall**

\*\* Accuracy \*\* measures how often the classifier makes the correct prediction. It’s the ratio of the number of correct predictions to the total number of predictions (the number of test data points).

\*\* Precision \*\* tells us what proportion of messages we classified as spam, actually were spam. It is a ratio of true positives(words classified as spam, and which are actually spam) to all positives(all words classified as spam, irrespective of whether that was the correct classificatio), in other words it is the ratio of

[True Positives/(True Positives + False Positives)]

\*\* Recall (sensitivity)\*\* tells us what proportion of messages that actually were spam were classified by us as spam. It is a ratio of true positives(words classified as spam, and which are actually spam) to all the words that were actually spam, in other words it is the ratio of

[True Positives/(True Positives + False Negatives)]

For classification problems that are skewed in their classification distributions like in our case where we have

* a total of 609 records with
* 180 individuals diagonised with ASD and
* 429 individuals not diagonised with ASD

accuracy by itself is not a very good metric. Thus, in this case precision and recall come in very handy. These two metrics can be combined to get the F1 score, which is weighted average(harmonic mean) of the precision and recall scores. This score can range from 0 to 1, with 1 being the best possible F1 score(we take the harmonic mean as we are dealing with ratios).

### AUC Score:

AUC is the percentage of the ROC plot that is underneath the curve