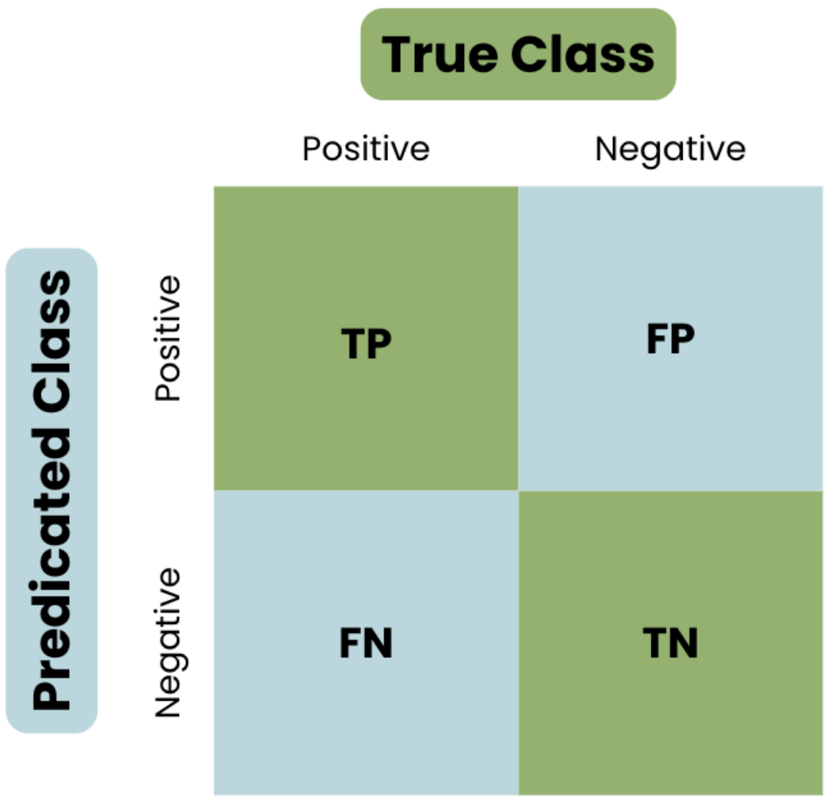
Evaluation and validation



Accuracy :

$$Acc = \frac{TP + TN}{TP + FP + FN + TN}$$

precision:

$$perc = \frac{TP}{TP + FP}$$

ideally, it should be one (the nominator = dominator) so the false positives will decrease it

Recall

$$recall = \frac{TP}{TP + FN}$$

again it should be one so make sure the false negatives are a little

---

F1

we want a metric that combines both precision and recall, taking the arithmetic mean is not enough here. so we use harmonic mean where one metric poor performance will highly affect the value (because we are dividing over 1/x so if x is small the value we are dividing by will be larger decreasing the final score)

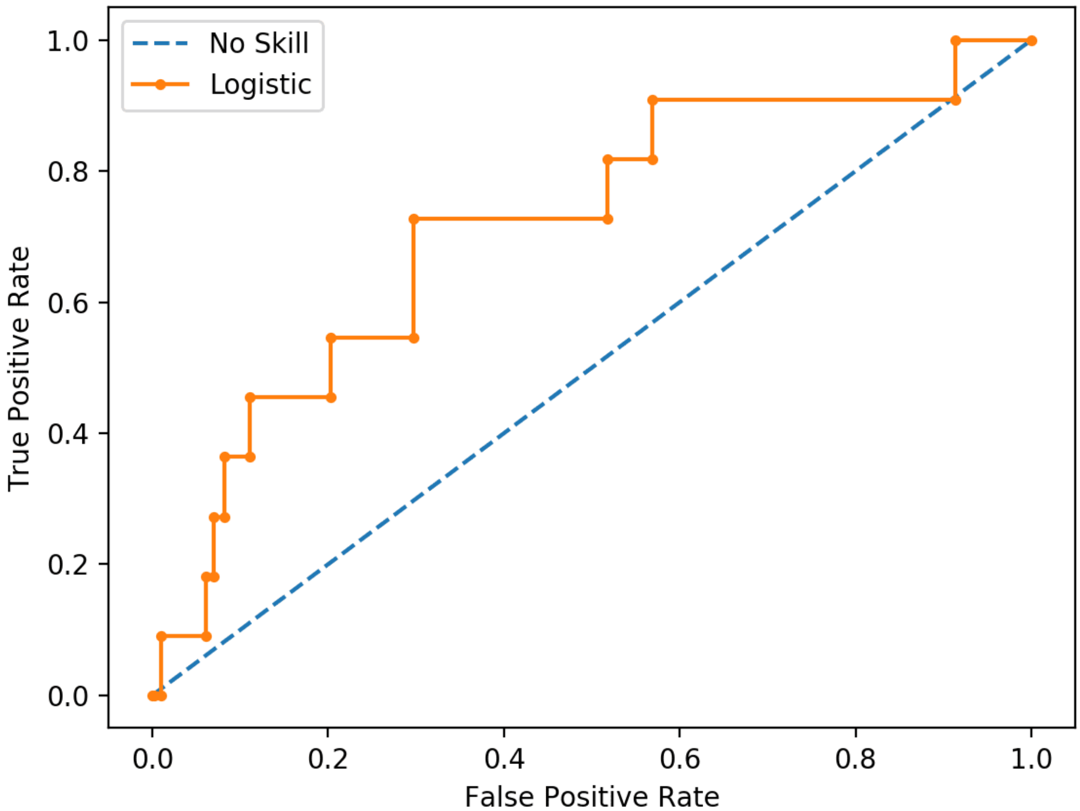
$$F1 = \frac{2}{\frac{1}{per} \frac{1}{recall}}$$

simplifying the equation:

$$F1 = 2 * \frac{perc * recall}{perc + recall}$$

ROC-AUC

The ROC (Receiver Operating Characteristic) curve plots the rate of TP against FP for different thresholds (like the threshold of logistic regression!)



Axes:

TP rate or sensitivity or recall: TP / TP+FN (sum the column in the confusion matrix)

FP rate or specificity: TN / FP+TN

AUC: between 0 and 1. A higher AUC indicates a better model performance.

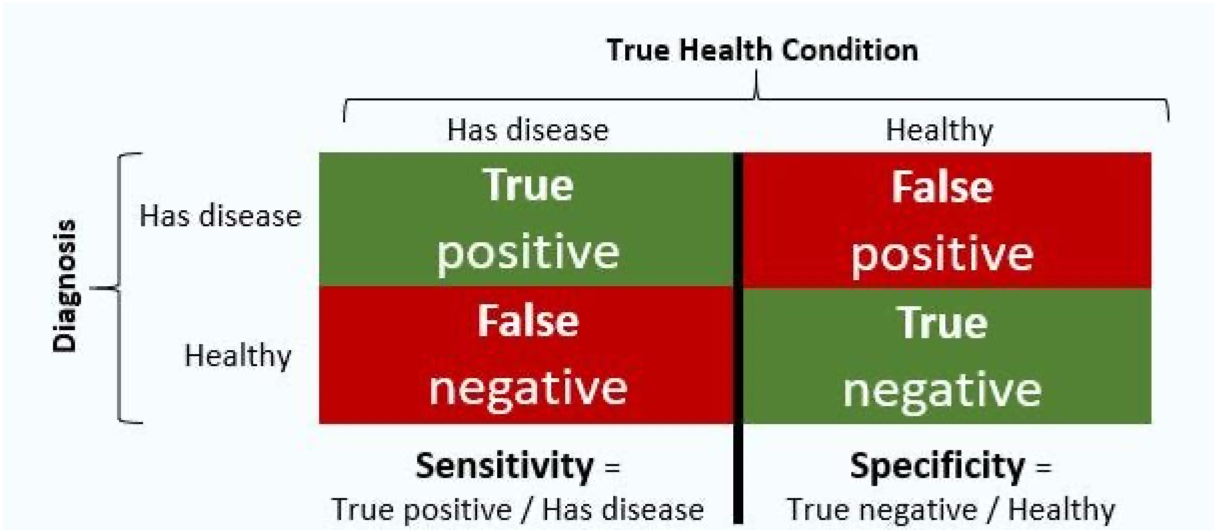
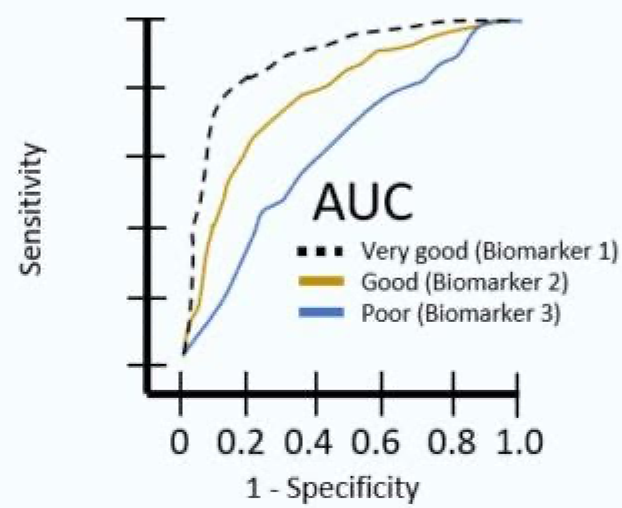


Figure 2. Calculation of senstivity and specificty.



helpful when working with unbalanced classes!