

## CONFUSION MATRIX

		Actual Class	
		p	n
Predicted Class	Y	True Positives	False Positives
	N	False Negatives	True Negatives
Totals:		P	N

True positive (TP): predicted to be positive and the actual value is also positive

False positive (FP): predicted to be positive but the actual value is negative

True negative (TN): predicted to be negative and the actual value is also negative

False negative (FN): predicted to be negative but the actual value is positive

**Accuracy:** Of all the classes, how many you predicted right. Accuracy is simply the fraction of the total sample that is correctly identified. Probability of giving correct predictions out of all predictions.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / \text{total}$$

**Precision:** Out of all the classes we have predicted as positive, how many are actually positive. Probability of getting positive predictions as correct out of total positive predictions.

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

Precision increases with increase in threshold value and decrease with decrease in threshold. With increase in threshold value, model is being more restrictive when classifying 1's which means False Positives errors would be less and True Positives will be more. With decrease in threshold value, model is being more restrictive when classifying 0's which means False Positives errors will be more and True Positives will be less.

**Recall/Sensitivity/True Positive Rate:** Out of all the actual positive classes, how many we predicted correctly. Sensitivity is the true positive rate (proportion of positives that are correctly identified).

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

Recall increases with decrease in threshold value and decrease with increase in threshold. With decrease in threshold value, model is being more restrictive when classifying 0's which means False Negative errors would be less and True Positive will be more. With increase in threshold value, model is being more restrictive when classifying 1's which means False Negatives errors will be more and True Positives will be less.

**Specificity/True Negative Rate:** Out of all the actual negative classes/values, how many we predicted correctly. Specificity is the true negative rate (proportion of negatives that are correctly identified).

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$$

**F measure:** It is difficult to compare two models with low precision and high recall or vice versa. So to make them comparable, we use F-Score. F-score helps to measure Recall and Precision at the same time. It uses Harmonic Mean in place of Arithmetic Mean by punishing the extreme values more. It combines precision and recall as a measure of effectiveness of classification. Lies between 0 and 1. Higher the F-score better the model. The F1 score combines precision and recall using the harmonic mean.

$$\text{F score} = 2 * \text{Recall} * \text{Precision} / (\text{Recall} + \text{Precision})$$

### **Problem with Confusion Matrix**

One of the problems with the confusion matrix is that all the values that are populated is based on an arbitrary choice of the threshold. What if we could evaluate multiple thresholds and see their impact. This can be done by usage of AUC ROC curve. These Curves also help in deciding the threshold when you have unbalanced data with you as it would show results at a number of thresholds and you can finalize the threshold value based on the graph.