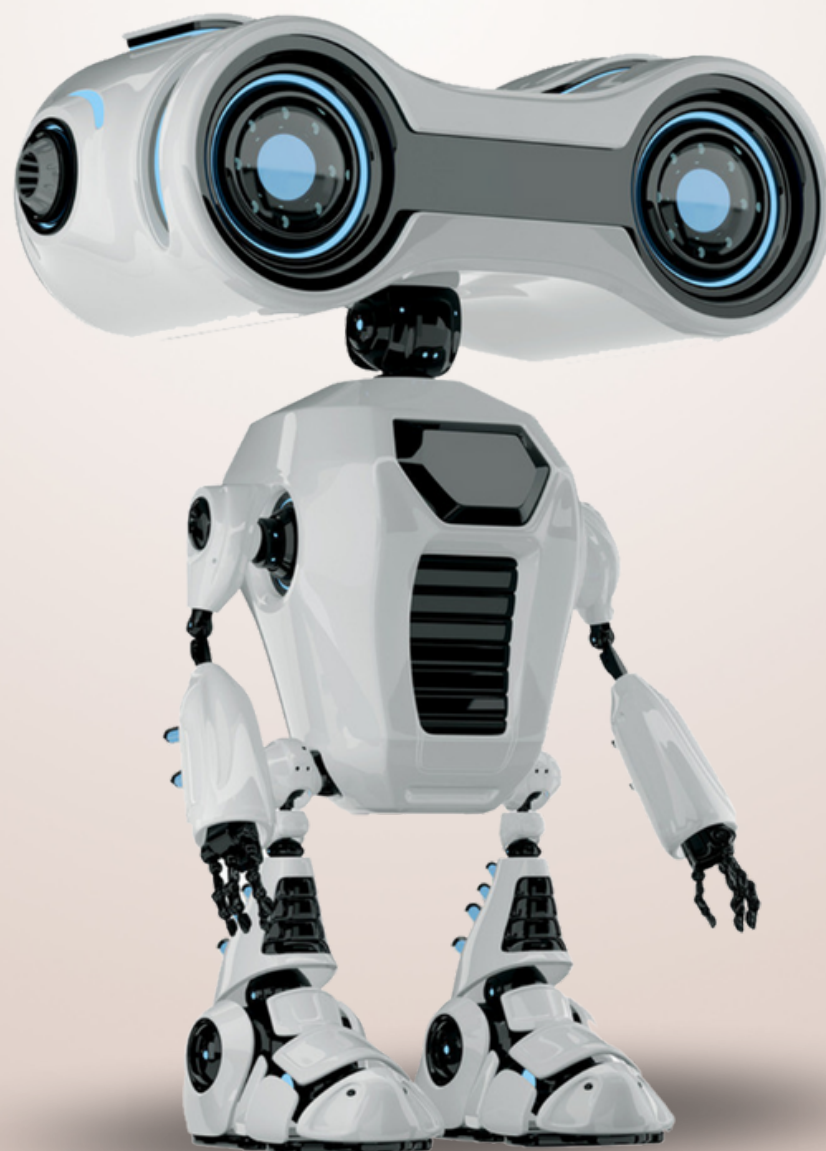


# MACHINE LEARNING TIPS CHEATSHEET

PART- I



# Metrics

Given a set of data points  $\{x^{(1)}, \dots, x^{(m)}\}$ , where each  $x^{(i)}$  has  $n$  features, associated to a set of outcomes  $\{y^{(1)}, \dots, y^{(m)}\}$ , we want to assess a given classifier that learns how to predict  $y$  from  $x$ .

# Classification

In a context of a binary classification, here are the main metrics that are important to track to assess the performance of the model.

# Confusion matrix –

The confusion matrix is used to have a more complete picture when assessing the performance of a model. It is defined as follows:

		Predicted class	
		+	-
Actual class	+	<b>TP</b> True Positives	<b>FN</b> False Negatives Type II error
	-	<b>FP</b> False Positives Type I error	<b>TN</b> True Negatives

# Main metrics –

The following metrics are commonly used to assess the performance of classification models:

Metric	Formula	Interpretation
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Overall performance of model
Precision	$\frac{TP}{TP + FP}$	How accurate the positive predictions are
Recall Sensitivity	$\frac{TP}{TP + FN}$	Coverage of actual positive sample
Specificity	$\frac{TN}{TN + FP}$	Coverage of actual negative sample
F1 score	$\frac{2TP}{2TP + FP + FN}$	Hybrid metric useful for unbalanced classes

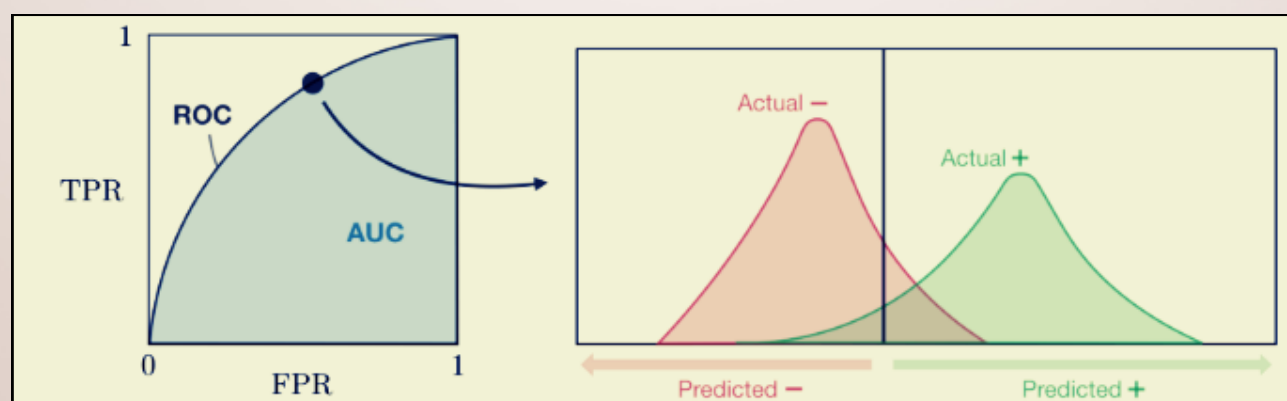
# ROC-

The receiver operating curve, also noted ROC, is the plot of TPR versus FPR by varying the threshold. These metrics are are summed up in the table below:

Metric	Formula	Equivalent
True Positive Rate TPR	$\frac{TP}{TP + FN}$	Recall, sensitivity
False Positive Rate FPR	$\frac{FP}{TN + FP}$	1-specificity

# AUC –

The area under the receiving operating curve, also noted AUC or AUROC, is the area below the ROC as shown in the following figure:



# Build your career story with 1stepGrow Academy

Follow 1stepGrow Academy

Save the Post

