AUC-ROC

Sunday 14 April 2024 11:07 AM

Why Roc Curve?

ROC curve is commonly used to compare the performance of models. It is usually used in binary classification, but it can also be used in multiclass classification using averaging methods.

→ The AUC-ROC curve, or Area Under the Receiver Operating Characteristic curve, is a graphical representation of the performance of a binary classification model at various classification thresholds. It is commonly used in machine learning to assess the ability of a model to distinguish between two classes

→ In a Binary classification two types of errors are possible
- Type I
- Type II

in some situations one type of error might be more important than the other
- a way to reduce a certain type of error is by adjusting the threshold value

→ But what theroshold value is right?
- ROC curve allows us to take an educated decision.

→ True Positive Rate/Recall [Benefit]
- $\frac{TP}{TP+FN}$ ↑ good
- Max(TPR) When FN=0
- Min(TPR) When TP=0

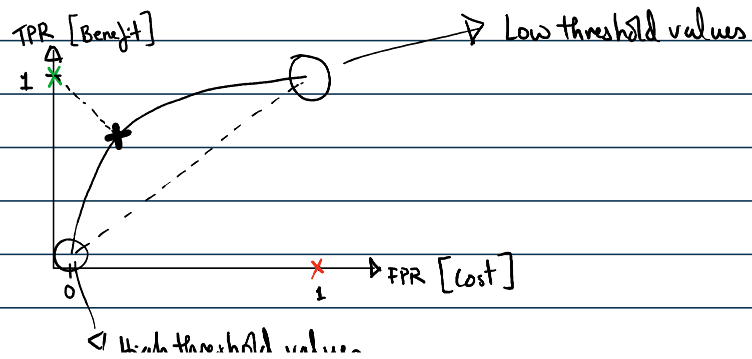
→ False positive rate (FPR) [Cost]
- $\frac{FP}{FP+TN}$ ↓ good
- Max(FPR) when TN=0
- Min(FPR) When FP=0

FPR
(False positive Rate)
↳ how many negative instances are incorrectly classified as positive instances out of all negative instances.

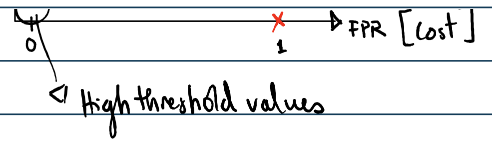
→ Ideal case TPR 100% FPR 0% [When there are 0 misclassifications]

100	0
0	100

→ ROC curve [plots TPR vs FPR at different threshold values]

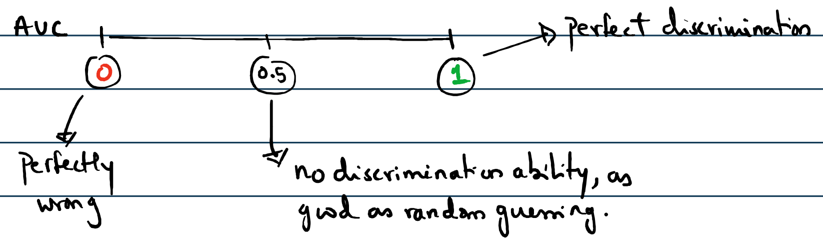


from sklearn.metrics import roc_curve



→ **AUC-ROC**

- measures the entire 2D area under the ROC curve from (0,0) to (1,1).
- It provides an aggregate measure of performance across all possible classification thresholds.
-



Imbalanced data? — Yes → AUPRC
Imbalanced data? — No → AUROC

Why? Incase of imbalanced data ROC overestimates the goodness of the model