Model performance measures

a. Confusion Matrix – A 2X2 tabular structure reflecting the performance of the model in four blocks

Confusion Matrix	Predicted Positive	Predicted Negative
Actual Positive	True Positive	False Negative
Actual Negative	False Positive	True Negative

a. Accuracy – How accurately / cleanly does the model classify the data points. Lesser the false predictions, more the accuracy

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

a. Sensitivity / Recall – How many of the actual True data points are identified as True data points by the model . Remember, False Negatives are those data points which should have been identified as True.

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

a. Specificity – How many of the actual Negative data points are identified as negative by the model

$$SPEC = \frac{TN}{TN + FP}$$

a. Precision – Among the points identified as Positive by the model, how many are really Positive

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

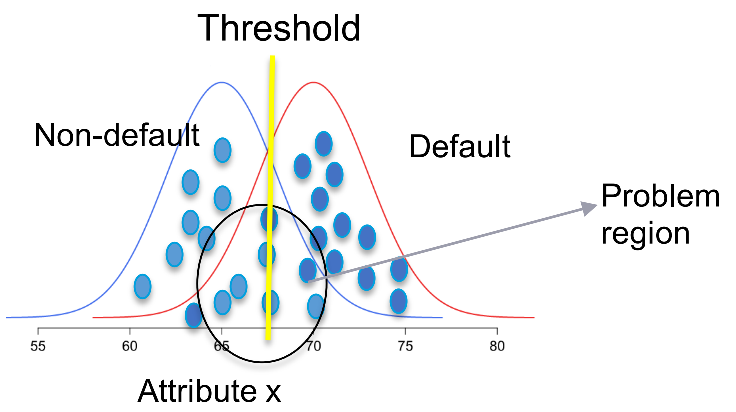
Model performance measures

Assume model is identifying defaulters. In this binary classification defaulter class is class of interest and labeled as +ive (positive - 1) class, other class is –ve(negative - 0)

1. True Positives - cases where the actual class of the data point and the predicted is same. For e.g. a defaulter (1) predicted as defaulter (1)
2. True Negatives – cases where the actual class was non-defaulter and the prediction also was non-defaulter
3. False Positives – cases where actual class was negative (0) but predicted as defaulter (1)
4. False Negatives – cases where the actual class was positive (1) but predicted as non-defaulter (0)
5. Ideal scenario will be when all positives are predicted as positives and all negatives are predicted as negatives

Model performance measures

6. In practical world this will never be the case. There will be some false positives and false negatives
7. Our objective will be to minimize both but the problem is, when we minimize one the other will increase and vice versa!
8. The problem is in the overlap region in the distributions



6. Objective will be to minimize one of the error types, either the false positive or false negative

Model performance measures

10. Minimize false negatives - if predicting a positive case as negative is going to be more detrimental for e.g. predicting a cancer patient (positive) as non-cancer (negative)
11. Minimize false positives – if predicting a negative as positive is going to be more detrimental for e.g. predicting a boss's mail as spam!
12. Accuracy – over all correct predictions from all the classes to total number of cases. Should rely on this metrics only when all classes are equally represented. Not reliable if class representation is lopsided as algorithms are biased towards over represented class
13. Precision - TP/ TP+ FP. When we focus on minimizing false negatives, TP will increase but along with it FP will also increase. How much increase in TP starts hurting (due to increase in FP) ?

Model performance measures

14. Recall – TP / TP + FN : when we reduce FN to increase TP, how much we gain ? Recall and precision will oppose each other. We want recall to be as close to 1 as possible without precision being too bad

14. To compare models, we use ROC AUC that gives us the optimal combination of these metrics

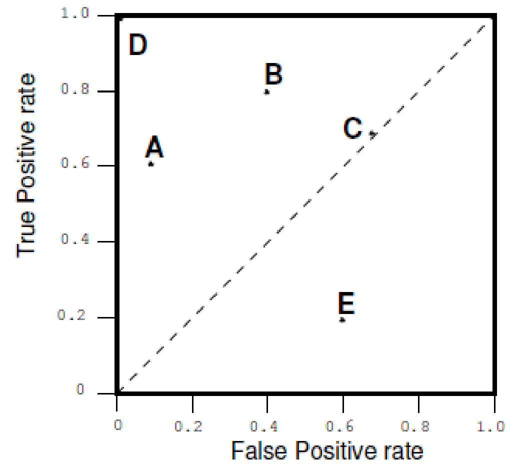
Receiver Operating Characteristics (ROC) Curve

A technique for visualizing classifier performance

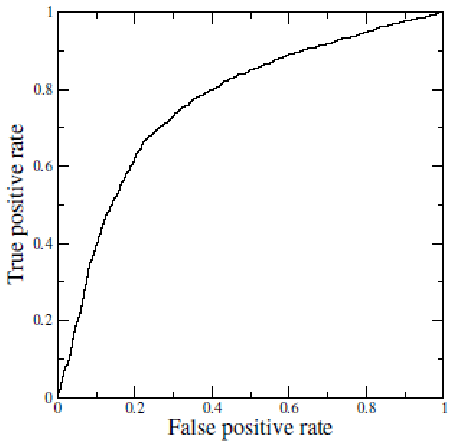
a. It is a graph between TP rate and FP rates
   I. TP rate = TP / total positive
   II. FP rate = FP / total negative
b. ROC graph is a trade off between benefits (TP) and costs (FP)
c. The point (0,1) represents perfect classified (e.g. D)
   I. TP = 1 and FP = 0

d. Classifiers very close to Y axis and lower (nearer to x axis) are conservative models and strict in classifying positives (low TP rate)

a. Classifiers on top right are liberal in classifying positives hence higher TP rate and FP rate



A basic ROC graph showing five discrete classifiers.



Ref:ROC_AUC.ipynb ,