AUC ROC Curve In Machine Learning

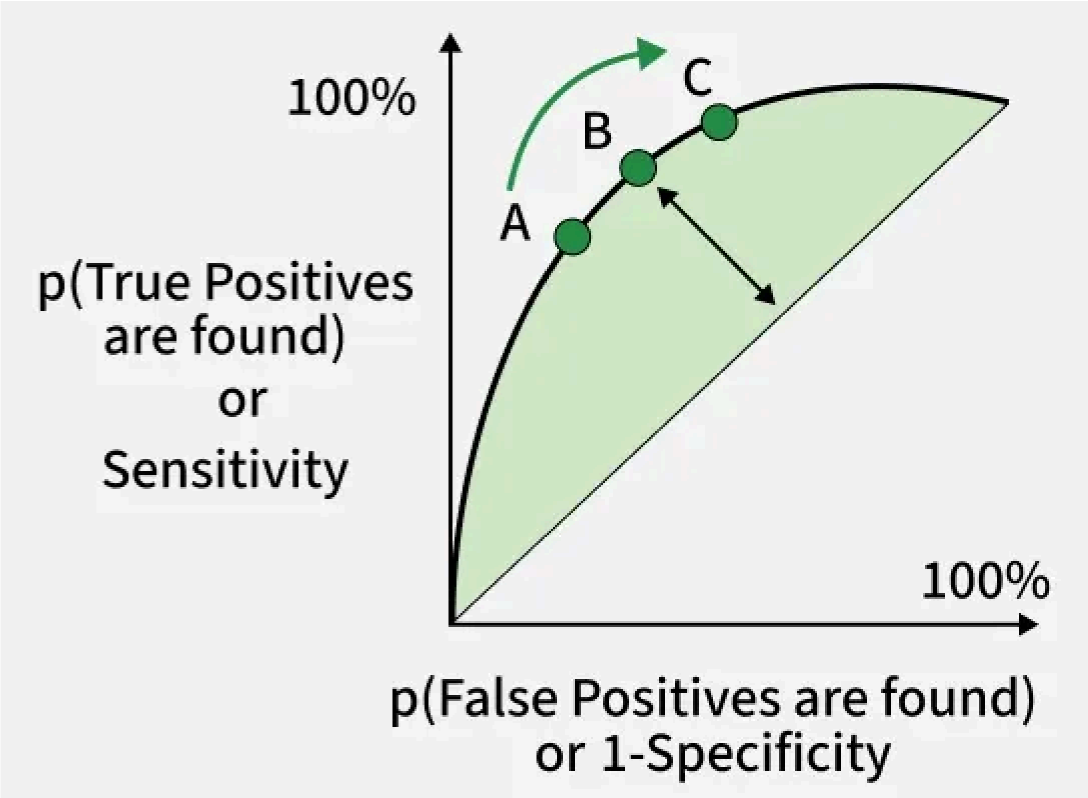
The larger the AUC is, the better the classifier is
ROC = Receiver Operator Characteristic
AUC = The area under the curve

AUC-ROC curve is a graph used to check how well a binary classification model works.

It helps us to understand how well the model separates the positive cases like people with a disease from the negative cases like people without the disease at different threshold level. It shows how good the model is at telling the difference between the two classes by plotting:

- **True Positive Rate (TPR) = Sensitivity:** how often the model **correctly** predicts the positive cases as Sensitivity or Recall.
  TPR = True positive/True Positive + False Positive
  if TPR = 1 = when the threshold is so low that every single sample is classified as obese, is 1
- **False Positive Rate (FPR):** how often the model **incorrectly** predicts a negative case as positive.
  FPR = False Positive/False Positive + True Negative
  if FPR = 1 = when the threshold is so low that every single sample that was not obese was incorrectly classified as obese
  false negative = how model incorrectly predict a positive case as negative case
- ***Specificity:** measures the proportion of actual negatives that the model correctly identifies. It is calculated as 1 - FPR.

The higher the curve the better the model is at making correct predictions.



All of these terms are derived from the confusion matrix table

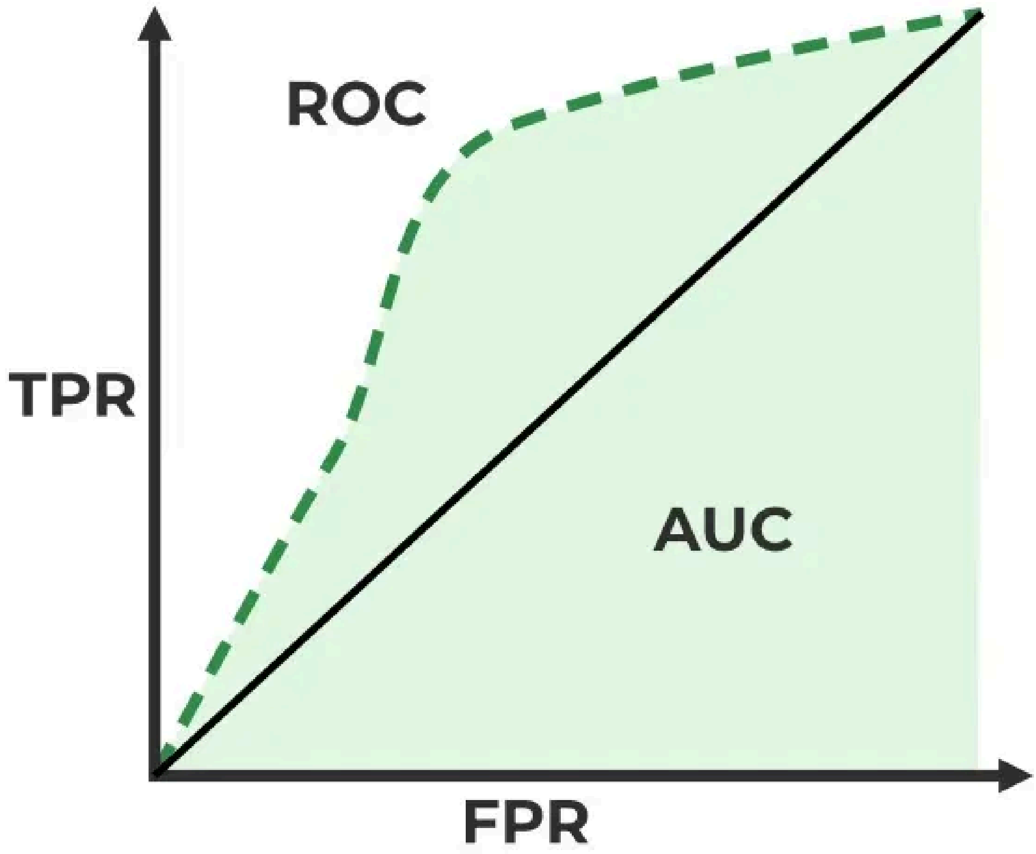
	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

- ROC curve = it plots TPR vs FPR at different thresholds. It represents the trade-off between the **sensitivity** and **specificity** of a classifier.
- AUC = measure the area under the ROC curve. A higher AUC value indicates better model performance as it suggests a greater ability to distinguishes between classes.
  -AUC = 1.0 = perfect performance
  -AUC = 0.5 = random guessing
  -AUC close to 0 = struggle to differentiate between the two classes

Example on how AUC-ROC work

AUC-ROC curve helps us understand how well a classification model distinguishes between the two classes. Imagine we have 6 data points and out of these:

- ***3 belong to the positive class:** Class 1 for people who have a disease.
- ***3 belong to the negative class:** Class 0 for people who don't have disease.



Now the model will give each data point a predicted probability of belonging to Class 1. The AUC measures the model's ability to assign higher predicted probabilities to the positive class than to the negative class. Here's how it work:

1. **Randomly choose a pair**: Pick one data point from the positive class (Class 1) and one from the negative class (Class 0).
2. **Check if the positive point has a higher predicted probability**: If the model assigns a higher probability to the positive data point than to the negative one for correct ranking.
3. ***Repeat for all pairs**: We do this for all possible pairs of positive and negative examples.

When?

-The dataset is balanced and the model needs to be evaluated across all thresholds.
-False positives and false negatives are of similar importance.

In cases of highly imbalanced datasets AUC-ROC might give overly optimistic results. In such cases the Precision-Recall Curve is more suitable focusing on the positive class.