- Evaluation Metrics for Classifiers

→ Evaluation Metrics

. After we build our model, we need to know how much our classifier is performing in terms of metrics that matter to our problem

. This is why we have several metrics that will tell us how good or bad our classifier is from different aspects.

1) Accuracy

. Metric that helps us identify the proportion of correct predictions the model has made compared to the total predictions carried

$$\text{Accuracy} = \frac{\text{Number of correctly predicted labels}}{\text{Total number of}}$$

2) Confusion Matrix

. Metric that helps us get an idea about what our classifier is good at.

Nb of predicted positive and are positive

Nb of predicted positive but are negative

TP	FP
FN	TN

Nb of predicted negative but are positive

Nb of predicted negative and are negative

. Frome those 4 components, we can calculate all the other metrics

↳ Example

y	y pred
0	1
1	1
0	0
1	1

$$\text{Accuracy} = \frac{\text{Nb of correctly predicted}}{\text{Total nb of examples}}$$

$$= 3/4 = 0.75$$

↳ So our accuracy is 75 %

TP	FP	Total Predicted Positive
FN	TN	Total Predicted Negative
Total Real Pos.	Total Real Neg.	

True pos. rate

$$TPR = \frac{TP}{\text{Total Real Pos.}}$$

↳ when its +, and how much we predict +

False pos. rate

$$FPR = \frac{FP}{\text{Total Real Neg.}}$$

↳ when its −, and how much do we predict +

False neg. rate

$$FNR = \frac{FN}{\text{Total Real pos.}}$$

↳ when its +, how much we predict −

True neg. rate

$$TNR = \frac{TN}{\text{Total Real neg.}}$$

↳ when its −, how much do we predict −

→ Other Metrics:

1) Precision (Positive Predictive Value)

It's the ratio of correctly predicted positive observations to the total predicted positives (How much we're correct when we're predicted positive)

$$Precision = \frac{TP}{\text{Total Predicted Pos.}}$$

↳ Interpretation: High precision means that the model rarely classifies negative instances as positive.
Precision is important in contexts where False Positives are costly; such as spam detection

2) Recall (Sensitivity or True Positive Rate)

It's the ratio of correctly predicted positive observations to all actual positives

$$Recall = \frac{TP}{\text{Total Real Pos.}}$$

↳ Interpretation: High Recall indicates that the model rarely misses positive cases
It's crucial in scenarios where False Negatives are costly, such as diagnosing diseases

3) F1 Score

- It's the weighted average for precision and recall

$$F_1 \text{ score} = \frac{2 \text{ Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

↳ The F1 score is usefull when we want to deal with imbalanced dataset where the nb of classes are not equally distributed High F1 score indicates a good balance

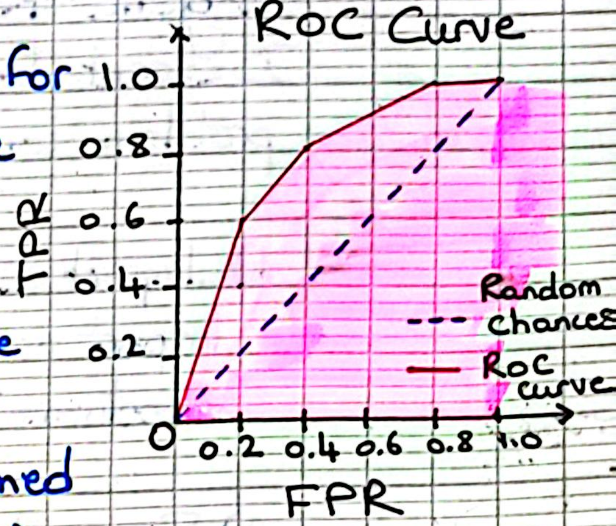
4) Roc Curve

. A Roc curve shows us the performance of a classifier while we adjust a threshold value to classify points between 0 and 1

. For example after obtaining the below values, we need to decide which are labeled as 1 and which as 0 according to a threshold T

Height	Gender	G_Pred	T=0.15	T=0.5	T=0.75
160	0	0.8	1	1	1
170	1	0.2	1	0	0
163	0	0.7	1	1	0
175	1	0.1	0	0	0
			TPR=0.5 FPR=1.0	TPR=0 FPR=1	TPR=0 FPR=0.5

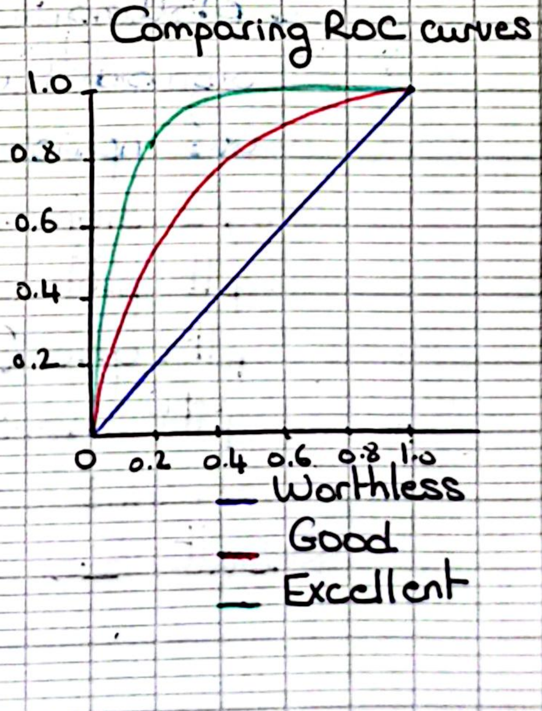
. For each of those classifiers (of different threshold T) we have a TPR and FPR



. After running the results for all possible thresholds, we can plot the TPR values with respect to FPR.
This is how we obtain the ROC curve.

. The highlited area is named AUC (Area Under the Curve)
A higher AUC indicates a better model
↳ AUC Interpretation
0.5: Model is no better than random guessing
0.7-0.8: Fair Performance
0.8-0.9: Good Performance
0.9-1.0: Excellant Performance

→ ROC Curve - Example



. Here's a comparision between 3 ROC curves

. We always aim to achieve high TPR versus a low FPR