Data Science and Artificial Intelligence

Machine Learning

Classification

Lecture No. 5

By- SIDDHARTH SABHARWAL SIR

Recap of Previous Lecture

- Topic
- Topic
- Topic
- Topic
- Topic

logistic Reg.

MLP

Kclass

Topics to be Covered

- Topic — dogistic Reg
- Topic — Auc
- Topic — Roc
- Topic — Recall / Precision
- Topic


DON'T
Complain
JUST
DOIT

REVISION

Logistic Regression :

To find β

- MLE
  - max $\Rightarrow \prod_{i=1}^{N} \left(\frac{1}{1+e^{-x_i\beta}}\right)^{y_i} \left(1-\frac{1}{1+e^{-x_i\beta}}\right)^{1-y_i}$
  - likelihood $\Rightarrow$ When class 1 $\Rightarrow \frac{1}{1+e^{-x_i\beta}}$
    " " 0 $\Rightarrow 1-\frac{1}{1+e^{-x_i\beta}}$

REVISION

Logistic Regression :

For K class, (K-1) Sigmoid

$$\log_e \frac{P(Y=1/X=x)}{P(Y=K/X=x)} = x\beta_1$$

$$P_i = \frac{e^{x\beta_i}}{1+\sum_{j=1}^{K-1} e^{x\beta_j}}$$

$$P_K = \frac{1}{1+\sum_{j=1}^{K-1} e^{x\beta_j}}$$

Logistic Regression : for multiclass case

- So for any new point find Probab of all classes, which ever is maxima will decide the class

So which will have highest likelihood ⇒

we have 4 classifier
how to find the classifier
with max likelihood

Concept ⇒

#Q. The following table gives the binary labels $(y^{(i)})$ for four points $\left(x_1^{(i)}, x_2^{(i)}\right)$ where $i = 1, 2, 3, 4$. Among the given options, which set of parameter values $\beta_0, \beta_1, \beta_2$ of a standard logistic regression model $p(x_i) = \frac{1}{1+e^{-(\beta_0+\beta_1 x+\beta_2 x)}}$ results in the highest likelihood for this data?

nice

(a) $\beta_0 = 0.5, \beta_1 = 1.0, \beta_2 = 2.0$

(b) $\beta_0 = -0.5, \beta_1 = -1.0, \beta_2 = 2.0$

(c) $\beta_0 = 0.5, \beta_1 = 1.0, \beta_2 = -2.0$

(d) $\beta_0 = -0.5, \beta_1 = 1.0, \beta_2 = 2.0$

x_1	x_2	y
0.4 ✓	−0.2 ✓	1
0.6	−0.5	1
−0.3	0.8	0
−0.7	0.5	0

$$\text{likelihood} = \prod_{i=1}^{4} \left(\frac{1}{1+e^{-x_i\beta}}\right)^{y_i} \left(1-\frac{1}{1+e^{-x_i\beta}}\right)^{1-y_i}$$

$$\text{likelihood} \Rightarrow \left(\frac{1}{1+e^{-x_1\beta}}\right) \cdot \frac{1}{1+e^{-x_2\beta}} \cdot \left(1-\frac{1}{1+e^{-x_3\beta}}\right) \cdot \left(1-\frac{1}{1+e^{-x_4\beta}}\right)$$

But it is very obvious that the classifier which classify the points correctly will have max likelihood.

Classifier option(a)

$x\beta \Rightarrow \beta_0 + \beta_1 x^1 + \beta_2 x^2$

$\Rightarrow .5 + x^1 + 2x^2$

for all 4 points

1) $x_1\beta \Rightarrow .5$ 2) .1 3) 1.8 4) .8

Class1

Classifier in option b.

$x\beta = -.5 - x^1 + 2x^2$

for all 4 points

1) -1.3 2) -2.1 3) 1.4 4) 1.2

Class0 — Class1

Best

Classifier option c

$x\beta = .5 + x^1 - 2x^2$

1) 1.3 4) -1.2

2) 2.1

3) -1.4

Classifier option

$-0.5 + x^1 + 2x^2$

1) -0.5 } Class 0 X
2) -0.9 }
3) $+0.8 \rightarrow$ Class 1 X
4) $-0.2 \rightarrow$ Class 0 ✓

- So the best classifier has max likelihood
- If in any Question two classifier are classifying perfectly then find likelihood

OR

Find $\sum_{i=1}^{N} y_i x_i \beta$ ($y_i = \pm 1$)

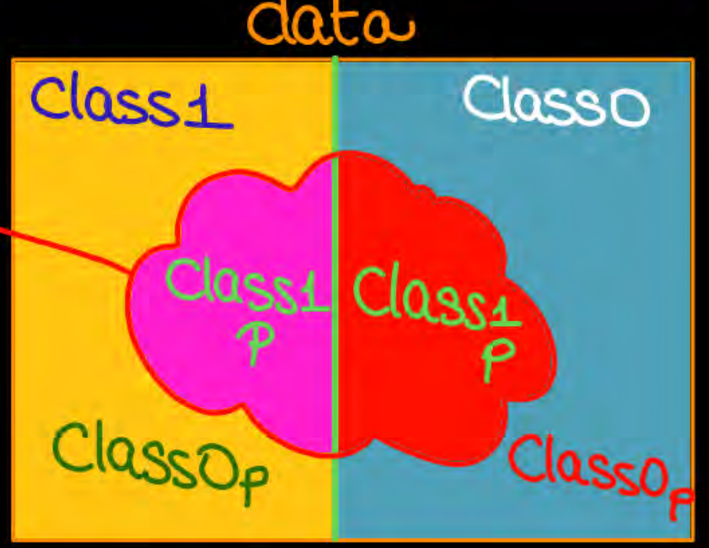
So classifier with $\max \sum_{i=1}^{N} y_i x_i \beta$ will have max likelihood.

Confusion Matrix

- we create Confusion matrix for checking the Performance of the classifier

Classifier → Class 1

Predicted Class	Actual Class 1	Actual Class 0
1	True Positive Pink	false Positive Red
0	false Negative Yellow	True Negative Blue



Linear Classification

Confusion Matrix

Linear Classification

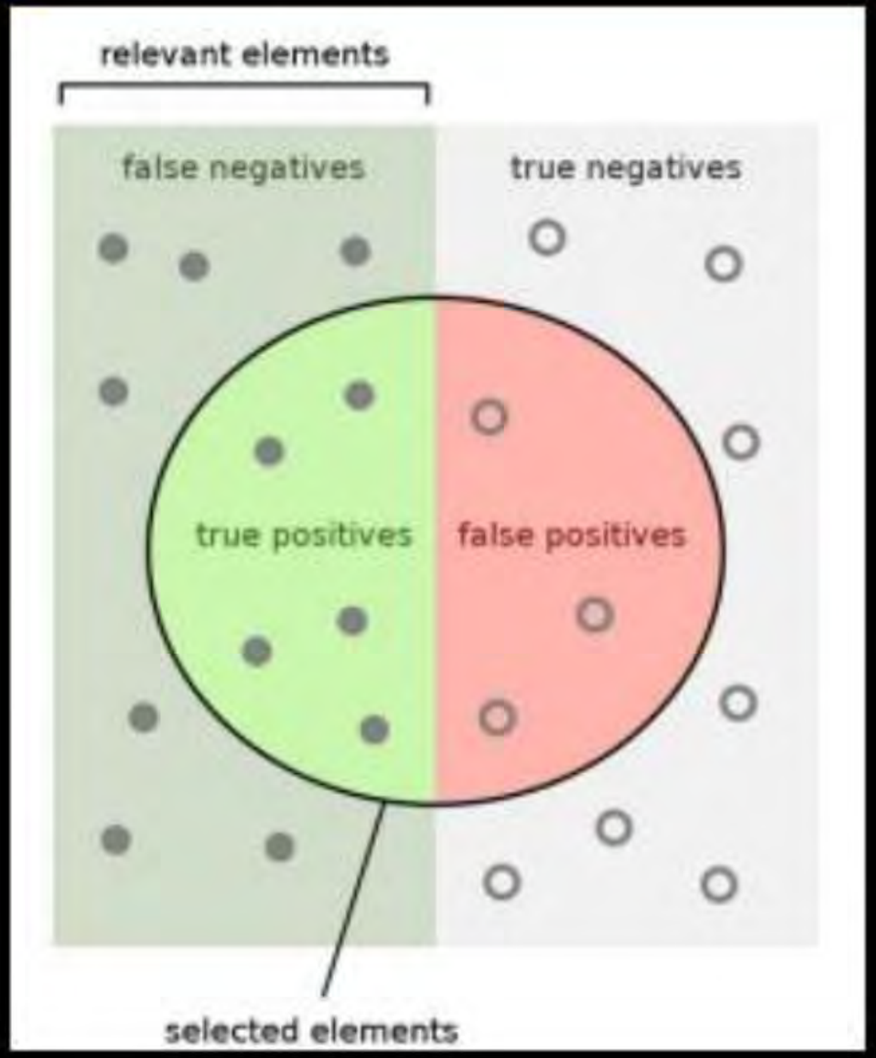
ROC ⇒ Reciever operating Ch. Curve.

What is ROC curve (receiver operating characteristic curve)

- A receiver operating characteristic curve, or ROC curve, is a graphical plot that illustrates the performance of a binary classifier model (can be used for multi class classification as well) at varying threshold values.
- The ROC curve is the plot of the true positive rate (TPR) against the false positive rate (FPR) at each threshold setting

Linear Classification

What is ROC curve (receiver operating characteristic curve)

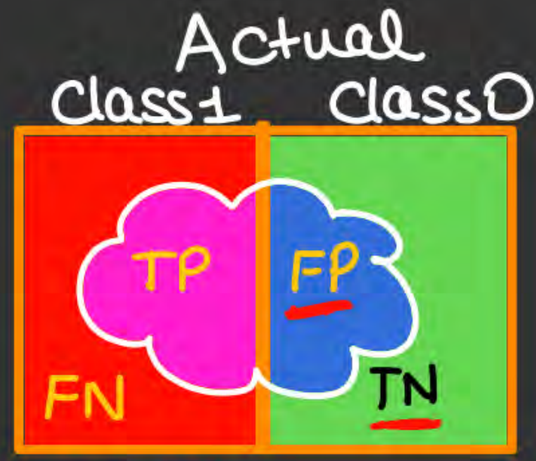


How many relevant items are selected? e.g. How many sick people are correctly identified as having the condition.

$$\text{Sensitivity} = \frac{}{}$$

How many negative selected elements are truly negative? e.g. How many healthy people are identified as not having the condition.

$$\text{Specificity} = \frac{}{}$$



Sensitivity ⇒ Accuracy to Predict Class 1

⇒ $\frac{TP}{\text{Actual P.}}$ ⇒ TPR

Sens ⇒ $\frac{TP}{\text{Actual P}}$

Speci ⇒ $\frac{TN}{\text{Actual N}}$ ⇒

Specificity ⇒ Accuracy to Predict Class 0

⇒ $\frac{TN}{\text{Actual N}}$ ⇒ $\left(1 - \frac{FP}{\text{Actual N}}\right) = (1 - FPR)$

⇓ $1 - \frac{FP}{\text{Actual N}}$

Linear Classification

What is ROC curve (receiver operating characteristic curve)

- Sensitivity is a measure of how well a test can identify true positives
- Specificity is a measure of how well a test can identify true negatives:

$$\text{sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}$$

$$\text{specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}}$$

Linear Classification

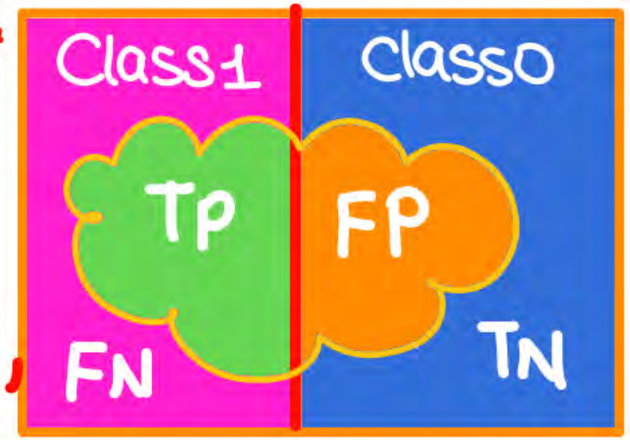
What is ROC curve (receiver operating characteristic curve)

- **What is TPR and FPR ?**

True Positive Rate (**TPR**) is a synonym for recall and is therefore defined as follows:

$$TPR = \frac{TP}{TP + FN}$$

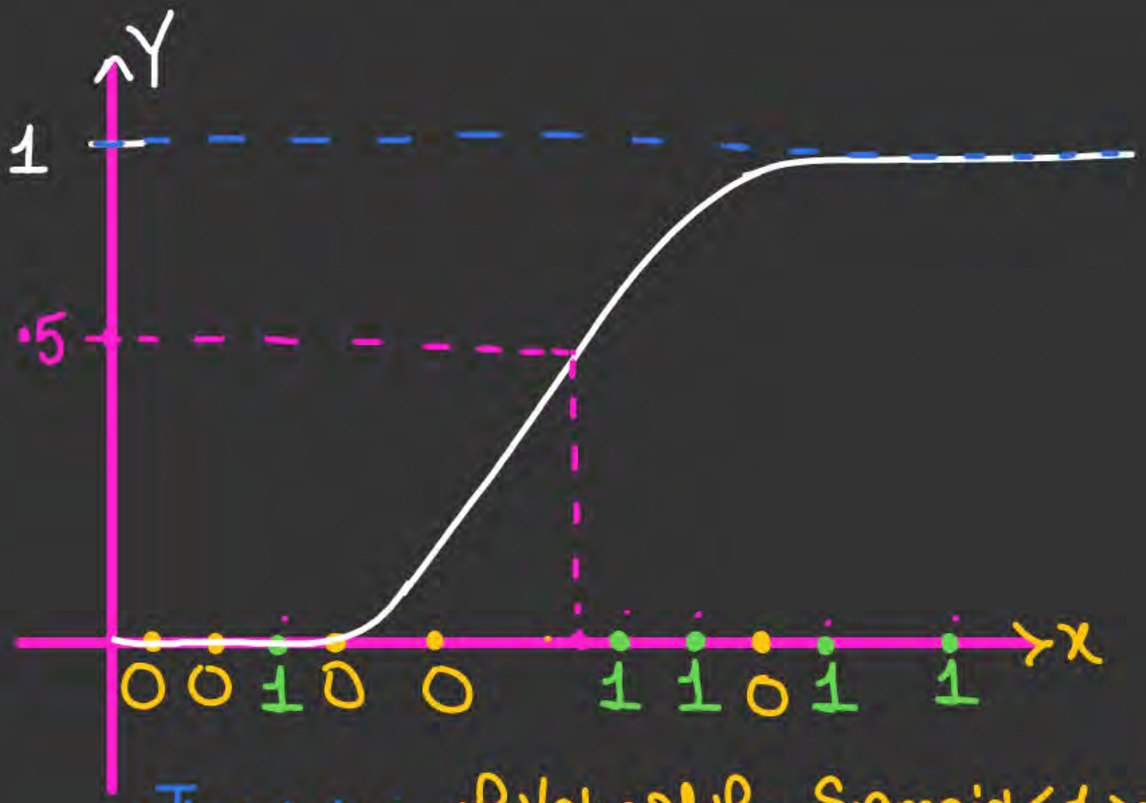
$TPR = \frac{TP}{TP+FN} = \frac{TP}{\text{Actual Class 1}}$



False Positive Rate (**FPR**) is defined as follows:

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

$FPR \Rightarrow \frac{FP}{FP+N} \Rightarrow \frac{FP}{\text{Actual Class 0}}$



- Threshold 1 if Value of the Sigmoid $< 1 \Rightarrow$ class 0
  $> 1 \Rightarrow$ class 1

So all predicted to 0

Predict

	1 Actual	0
1	0	0
0	5	5

Actual P
Predicted P

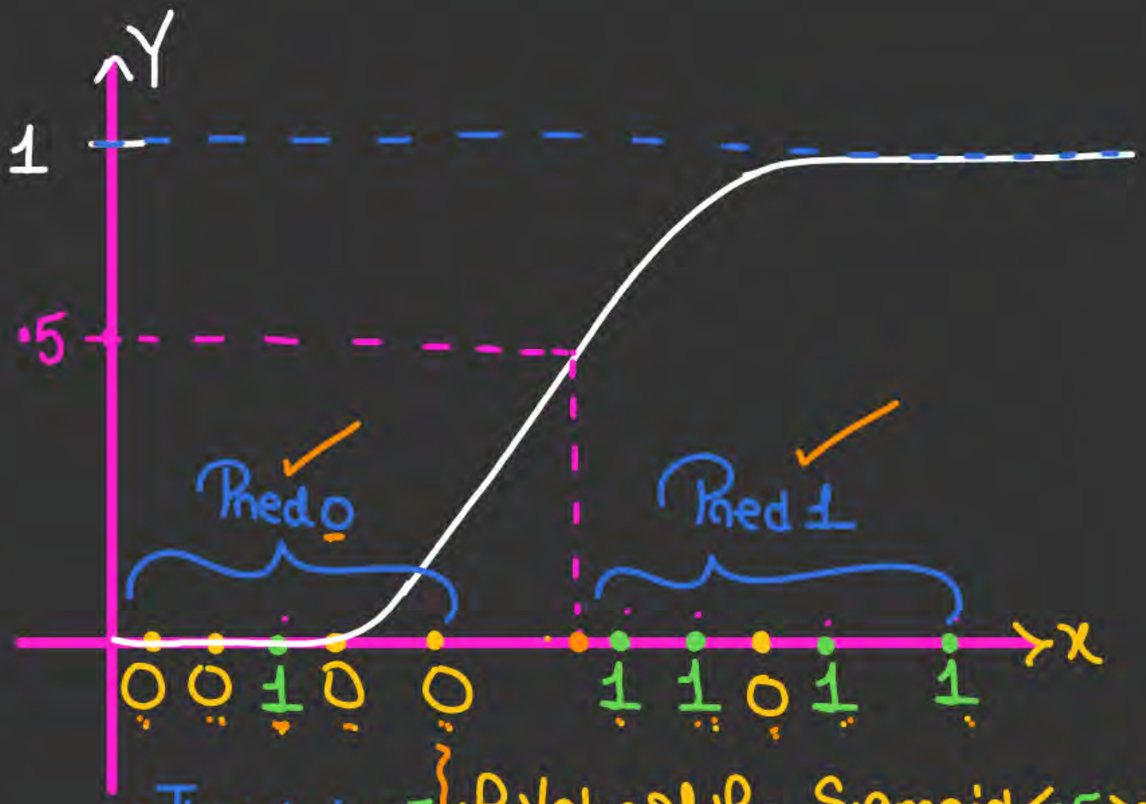
$$TPR \Rightarrow \frac{TP}{\text{Total Actual P}}$$

$$TPR \Rightarrow \frac{0}{5}$$

actual 0
Predicted P

$$FPR \Rightarrow \frac{FP}{\text{Total Actual N}}$$

$$FPR \Rightarrow \frac{0}{5}$$



• Threshold .5 if Value of the Sigmoid $\leq .5 \Rightarrow$ class 0
$> .5 \Rightarrow$ class 1

So all predicted to 0

Predict

	1 Actual	0
1	4	1
0	1	4

Actual P
Predicted P

$$TPR \Rightarrow \frac{TP}{\text{Total Actual P}}$$

$$TPR \Rightarrow 4/5$$

actual 0
Predicted P

$$FPR \Rightarrow \frac{FP}{\text{Total Actual N}}$$

$$FPR \Rightarrow 1/5$$

• Threshold .8 | If Value of the Sigmoid $\leq .8 \Rightarrow$ class 0
$> .8 \Rightarrow$ class 1

So all predicted to 0

Predict

	1 : Actual	0
1	2	1
0	3	4

Actual P
Predicted P

$$TPR \Rightarrow \frac{TP}{\text{Total Actual P}}$$

$$TPR \Rightarrow 2/5$$

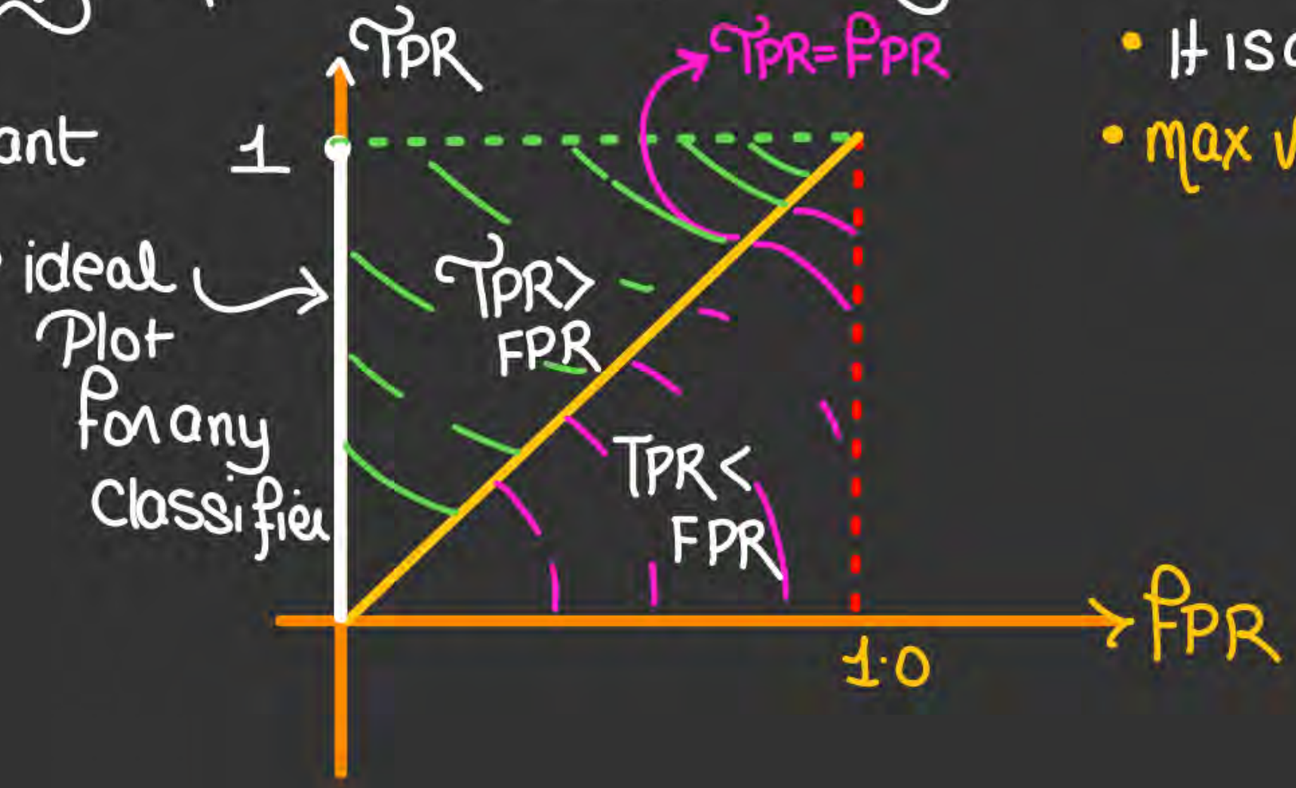
actual 0
Predicted P

$$FPR \Rightarrow \frac{FP}{\text{Total Actual N}}$$

$$FPR \Rightarrow 1/5$$

Roc ⇒ Reciever Operating Curve.

- Ideally we want FPR=0
-

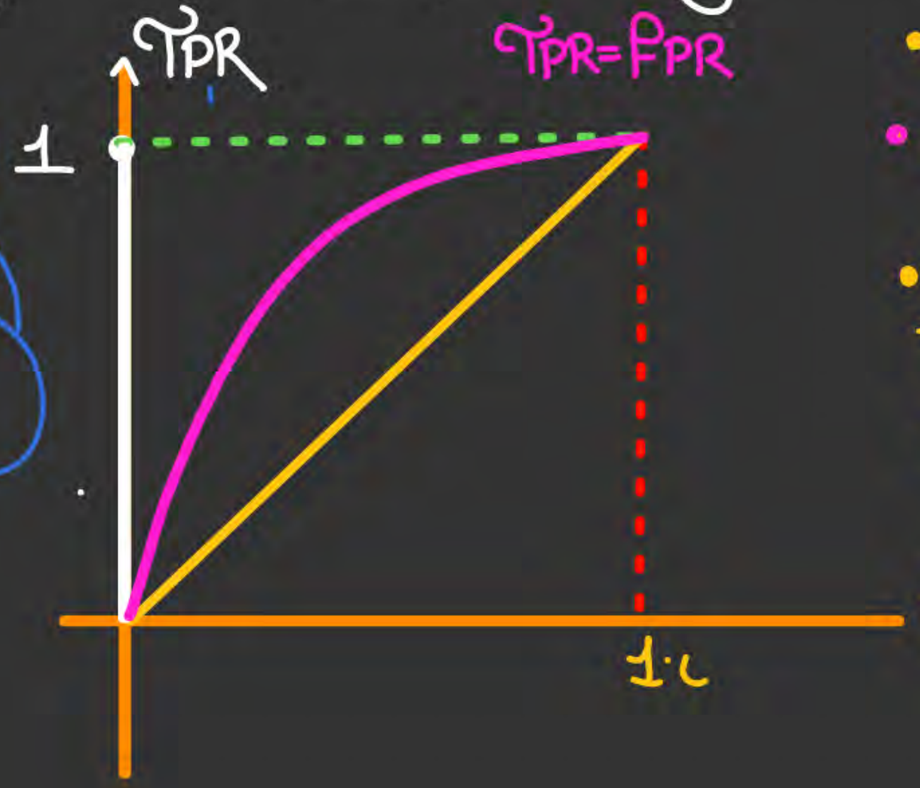


- ideal Plot for any Classifier

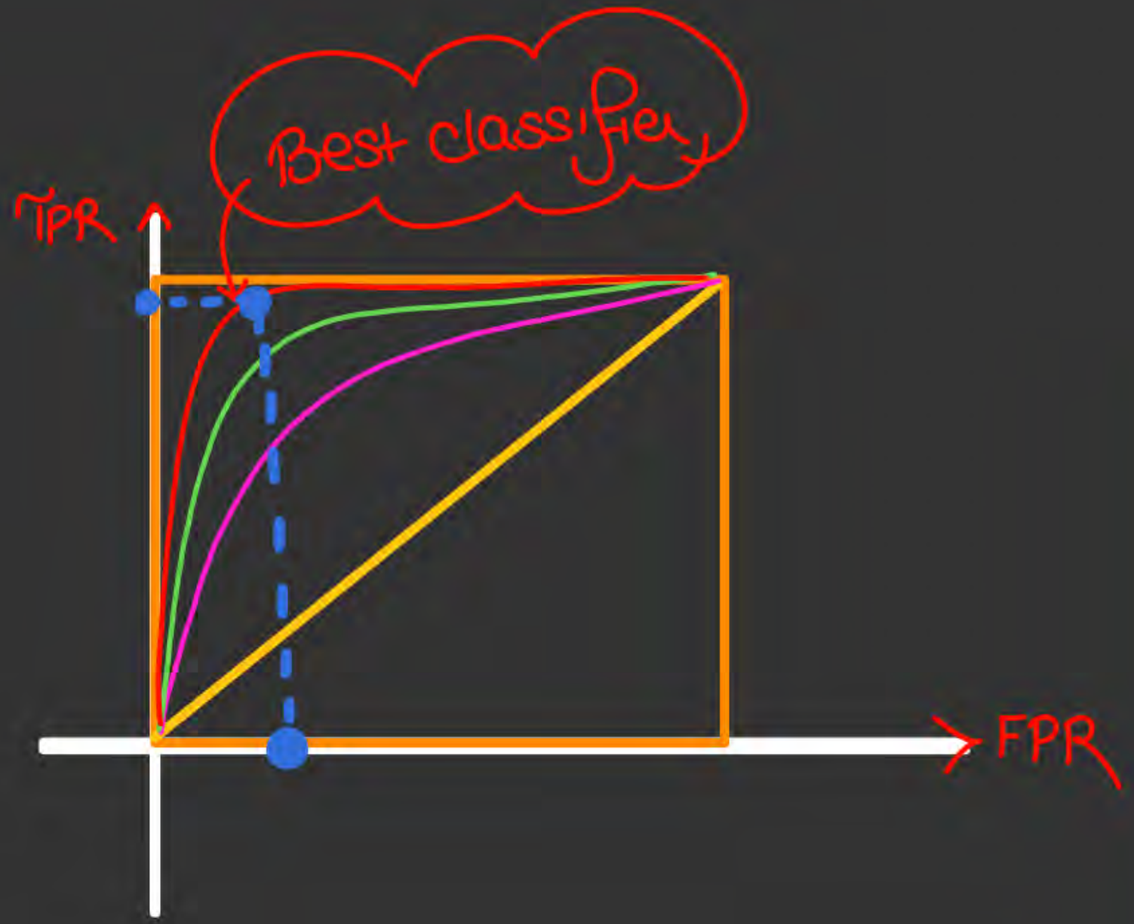
- It is a TPR, FPR curve
- max value TPR/FPR=1

Roc ⇒ Reciever Operating Curve.

In practical Case the plot shd Pass above the TPR = FPR line



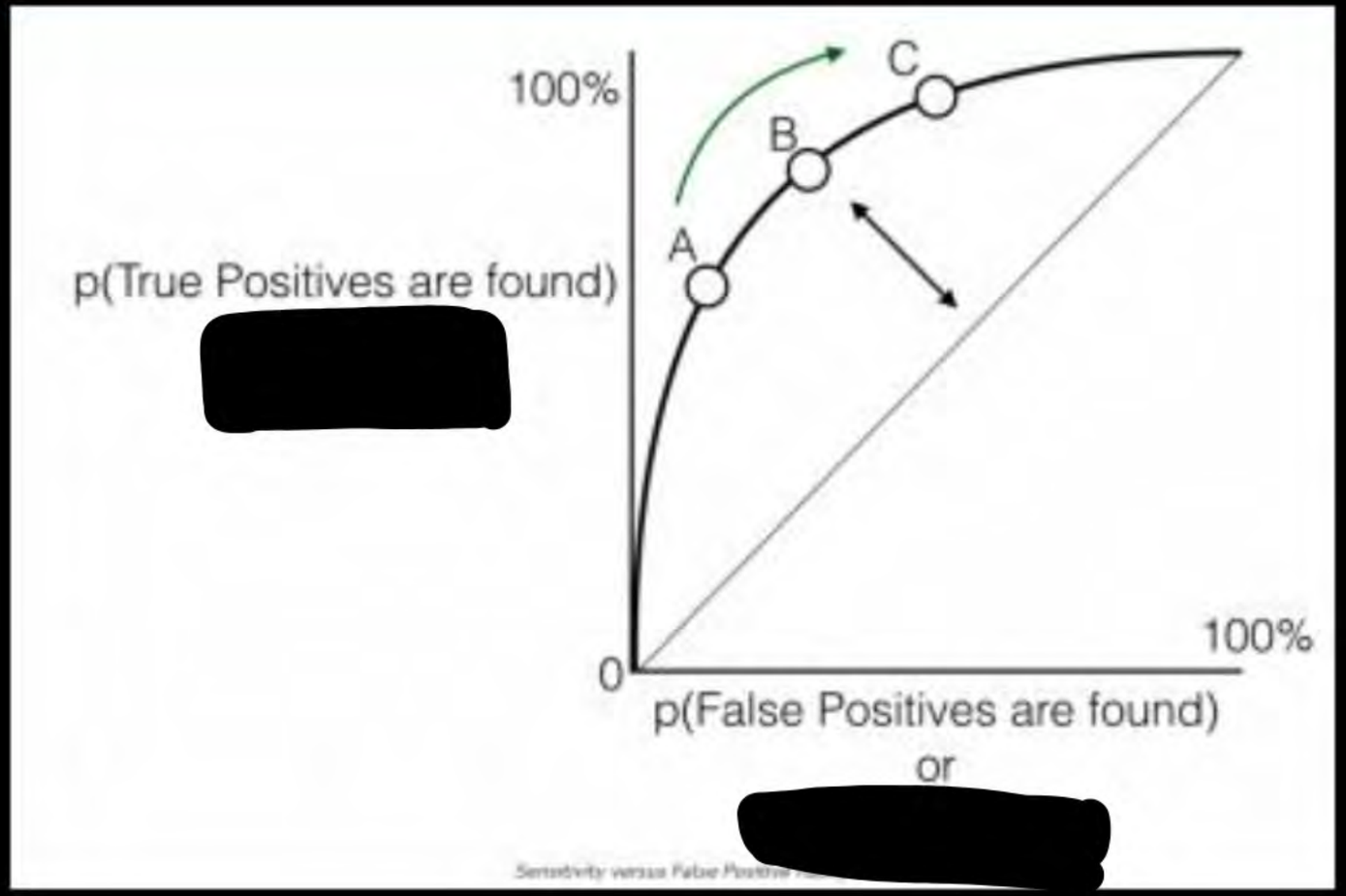
- It is a TPR, FPR curve
- max value TPR/FPR = 1
- So we get the plot of TPR/FPR by noting values of TPR and FPR for various thresholds of Classifier



- This ROC curve is used to Compare the Classifiers

Linear Classification

What is ROC curve (receiver operating characteristic curve) an example

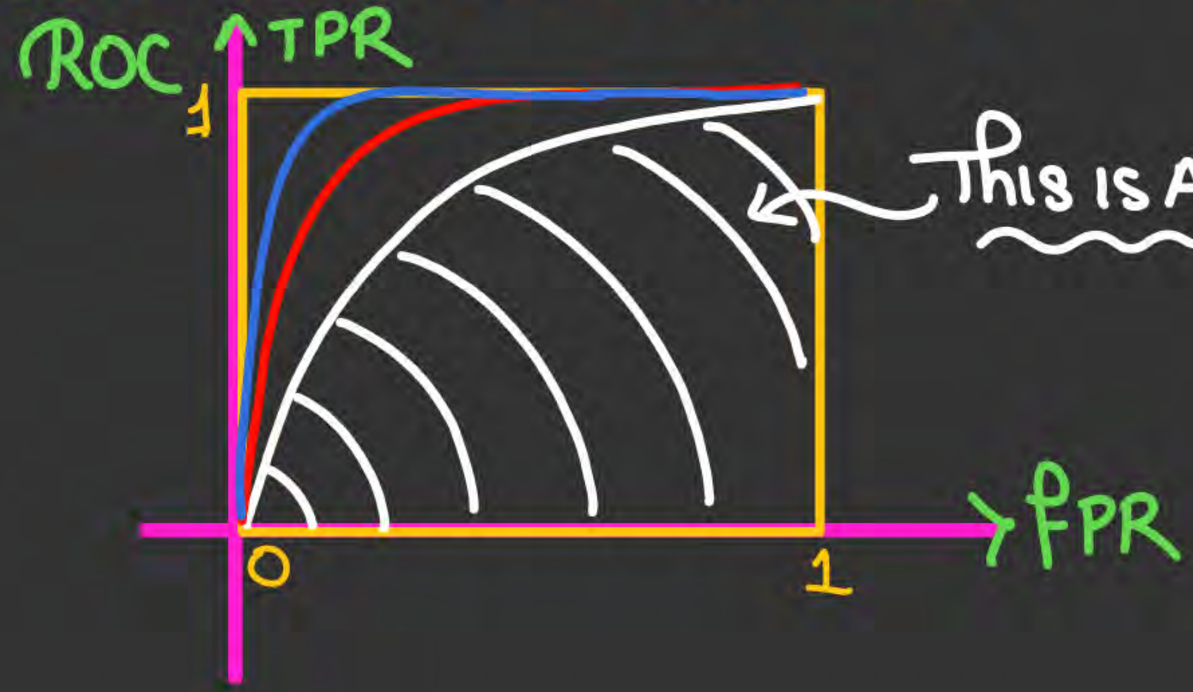


Linear Classification

What is AUC (Area under the curve)

- AUC stands for the Area Under the Curve, and the AUC curve represents the area under the ROC curve.
- It measures the overall performance of the binary classification model.
- The area will always lie between 0 and 1,
- A greater value of AUC denotes better model performance.
- Our main goal is to maximize this area in order to have the highest TPR and lowest FPR at the given threshold.
- The AUC measures the probability that the model will assign a randomly chosen positive instance a higher predicted probability compared to a randomly chosen negative instance.

So AUC → Area under the Curve.



, So the classifier which has max AUC is the best.

- AUC is always 0 to 1.

Linear Classification

What is AUC (Area under the curve)

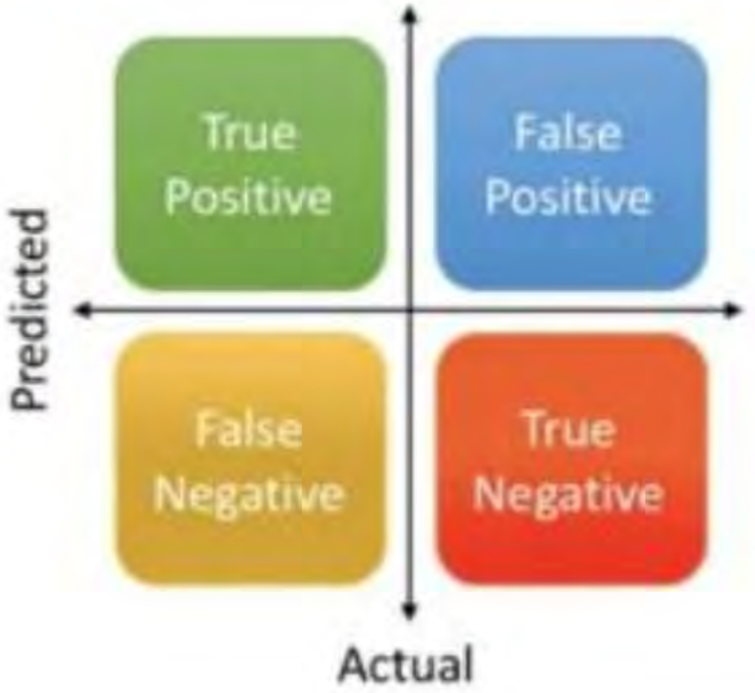
Linear Classification

What is Recall and Precision

$$\text{Precision} = \frac{\text{True Positive}}{\text{Actual Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

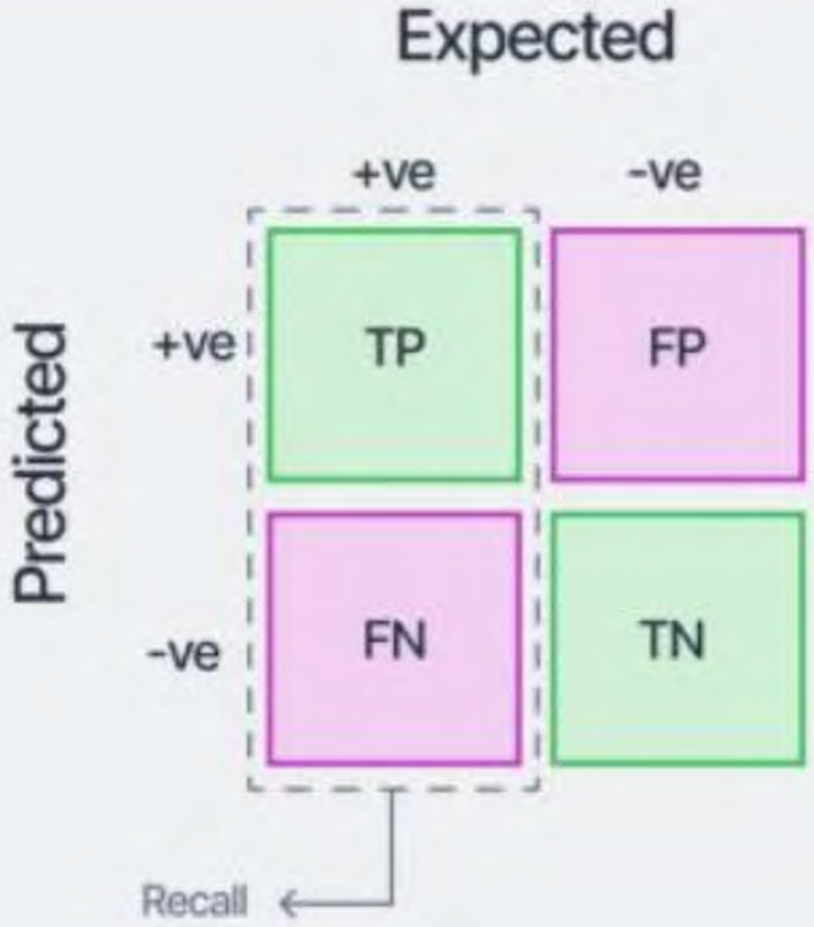
$$\text{Recall} = \frac{\text{True Positive}}{\text{Predicted Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive + False Negative}}$$

$$\text{Accuracy} = \frac{\text{True Positive + True Negative}}{\text{Total}}$$



Linear Classification

What is Recall and Precision



Linear Classification

What is Recall and Precision

Both precision and recall may be useful in cases where there is imbalanced data.

It may be valuable to prioritize one over the other in cases where the outcome of a false positive or false negative is costly.

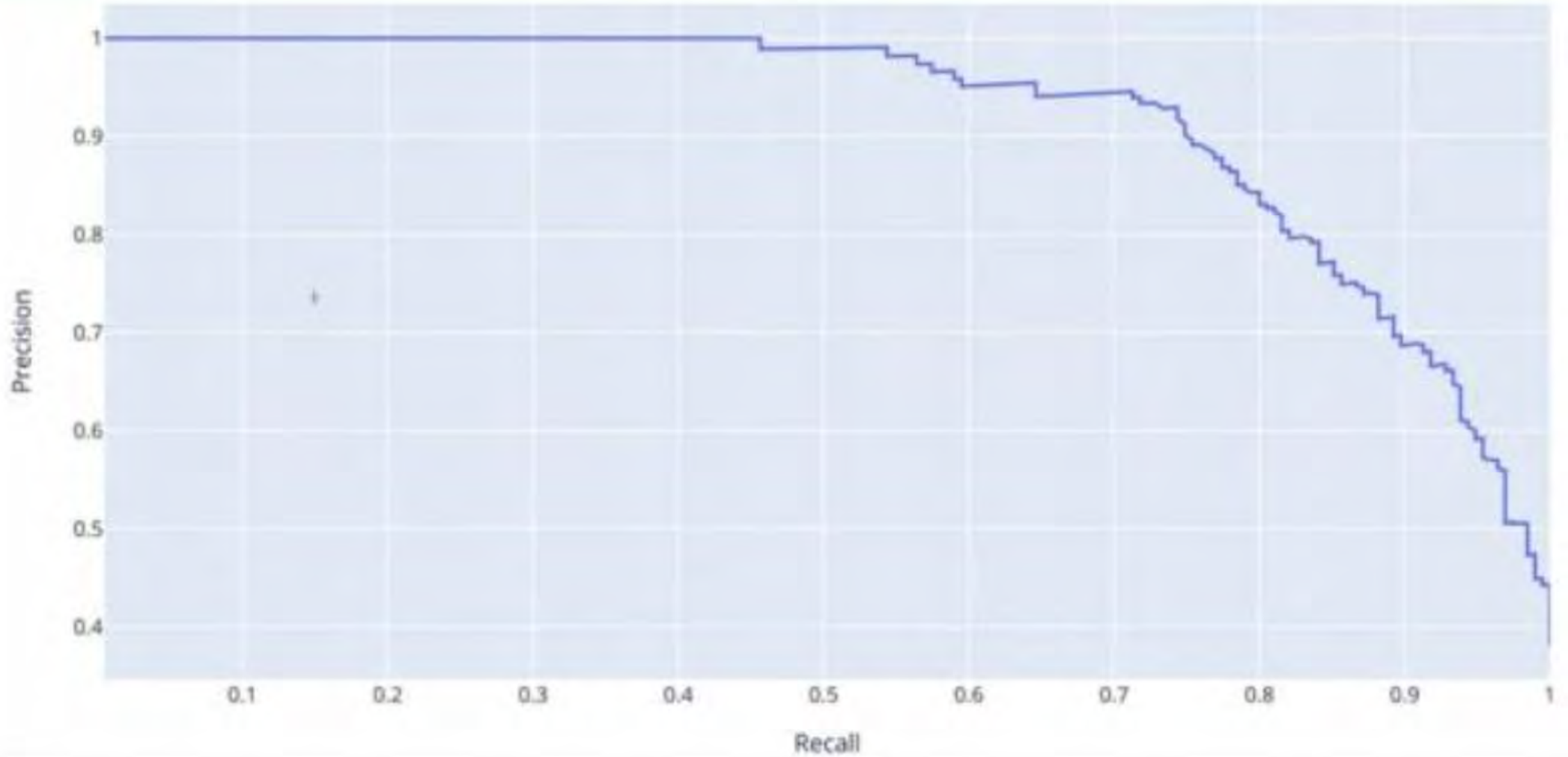
For example, in medical diagnosis, a false positive test can lead to unnecessary treatment and expenses.

In this situation, it is useful to value precision over recall. In other cases, the cost of a false negative is high.

For instance, the cost of a false negative in fraud detection is high, as failing to detect a fraudulent transaction can result in significant financial loss.

Linear Classification

What is Recall and Precision



Linear Classification

What is F-1 Score

In most problems, you could either give a higher priority to maximizing precision, or recall, depending upon the problem you are trying to solve. But in general, there is a simpler metric which takes into account both precision and recall, and therefore, you can aim to maximize this number to make your model better. This metric is known as F1-score, which is simply the harmonic mean of precision and recall.

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Linear Classification

Practise

The confusion matrix visualizes the ____ of a classifier by comparing the actual and predicted classes.

- Accuracy
- Stability
- Connectivity
- Comparativity

Linear Classification

Practise

From the above Table

n=200	Prediction=NO	Prediction = YES
Actual = NO	60	10
Actual = YES	5	125

- In reality, there are totally 135 accounts who have a balance more than $1000 and 70 accounts with balance - less than $1000
- In reality, there are totally 60 accounts who have a balance more than $1000 and 70 accounts with balance - less than $1000
- In reality, there are totally 125 accounts who have a balance more than $1000 and 10 accounts with balance - less than $1000
- In reality, there are totally 130 accounts who have a balance more than $1000 and 70 accounts with balance - less than $1000

Linear Classification

Practise

For the below confusion matrix, what is the recall?

	Not 5	5
Not 5	53272	1307
5	1077	4344

- ○ 0.7
- ○ 0.8
- ○ 0.9
- ○ 0.95

Linear Classification

What is F-1 Score

F1 score is:

- absolute mean of precision and recall
- harmonic mean of precision and recall
- squared mean of precision and recall

Linear Classification

What is F-1 Score

For the below confusion matrix, what is the F1 score?

	Not 5	5
Not 5	53272	1307
5	1077	4344

- ○ 0.72
- ○ 0.784
- ○ 0.82
- ○ 0.84

Linear Classification

What is F-1 Score

For a model to detect videos that are unsafe for kids, we need (safe video = postive class)

- ○ High precision, low recall
- ○ High recall, low precision

THANK - YOU