Data Science and Artificial Intelligence

Machine Learning

Classification

Lecture No. 4

By- SIDDHARTH SABHARWAL SIR

Recap of Previous Lecture

- **Topic** Logistic Regression
- **Topic** Logit
- **Topic** Sigmoid fxn.
- **Topic**
- **Topic**

Topics to be Covered

- Topic LogisticReg
- Topic ROC
- Topic AUC
- Topic Confusion matrix
- Topic

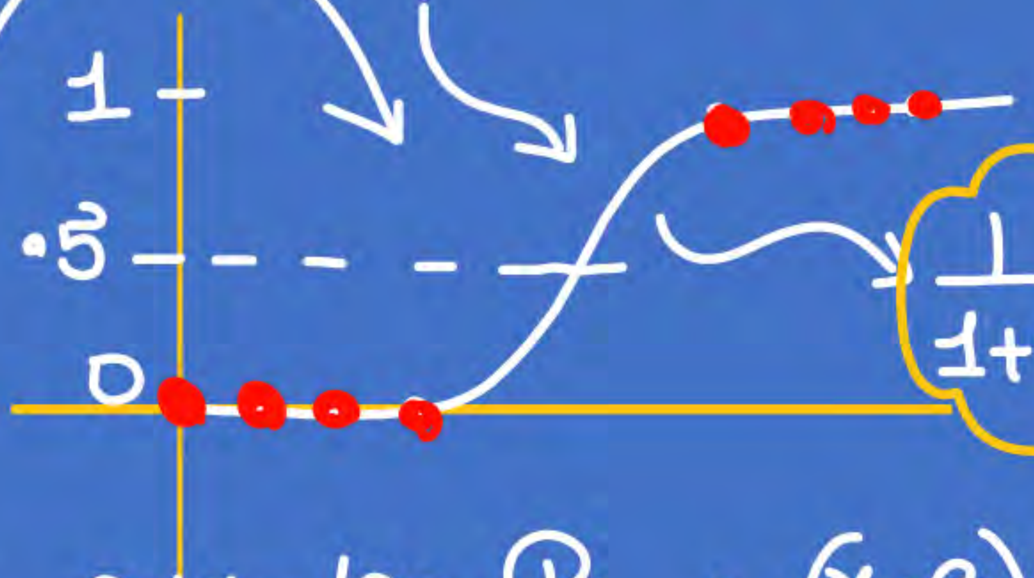

Inspiration comes from within yourself. One has to be positive. When you're positive, good things happen.
DEEP ROY
BRIAN TRACY
INTERNATIONAL

REVISION

Logistic Regression :

- Sigmoid fxn

S Curve



$$\frac{1}{1+e^{-x_i\beta}} \Rightarrow \text{How to find } \beta$$

$$\log_e \text{Odd} = \log_e \frac{P_1}{1-P_1} \Rightarrow (x_i\beta)$$

$$\Rightarrow \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2$$

REVISION

Logistic Regression :

done.

Linear Classification

Logistic Regression

8) ~~In continuation with question 7~~, let $x = 1$ if the server is wearing black shirt and $x = 0$ for servers wearing other colored shirts. We know that there are 270 observations with $x = 1$ and 340 observations with $x = 0$. The response variable is also an indicator variable given by $y = 1$ if the customer left a tip and $y = 0$ if the customer did not leave a tip. Use this data to fit a logistic regression model to compute the log-odds of leaving a tip depending on the color of the server's shirt..

→ $x=1$: Black shirt ; 270 observation, 140 leave tip.

$x=0$: white shirt ; 340 observation, 200 leave tip

Tip dena = Success ← $y=1$ if Customer left tip

$y=0$ if Customer donot leave tip

↳ failure

$\log_e \text{odd} = \beta_0 + \beta_1 x$

$x = 1$

$P_{success} = 140/270$

$P_{failure} = 130/270$

$x = 0$

$P_{success} = 200/340$

$P_{failure} = 140/340$

So $\log_e odd \Rightarrow \beta_1 x + \beta_0$

$\Rightarrow$ for $x = 1$

$\frac{P_{succ}}{P_{fail}}$

$\log_e \frac{140/270}{130/270} = \beta_1 + \beta_0$

$.0741 = \beta_1 + \beta_0$

$\underbrace{\beta_1 = -.282}$

for $x = 0$

$\log_e \frac{200/340}{140/340} = \beta_1 \times 0 + \beta_0$

$\beta_0 = .3566$

Linear Classification

Logistic Regression – Similarly we can have 2 D case also

ex ⇒ we will have 2 Dimension

$\Rightarrow \log_e odd = \beta_0 + \beta_1 x^1 + \beta_2 x^2$

we need 3 eq

x^1	x^2	Psuccess
1	5	.3
2	3	.2
5	6	.8

$$\left\{\begin{aligned} \beta_0 + \beta_1 + 5\beta_2 &= \log_e .3/.7 \\ \beta_0 + 2\beta_1 + 3\beta_2 &= \log_e .2/.8 \\ \beta_0 + 5\beta_1 + 6\beta_2 &= \log_e .8/.2 \end{aligned}\right\}$$

$\beta_0 = -3.72$

$\beta_1 = .43$

$\beta_2 = .48$

Logistic Regression

What type of dependent variable is suitable for logistic regression?

Y

A) Continuous variable
B) Categorical variable with multiple categories
C) Binary or dichotomous variable
D) Ordinal variable

Logistic Regression

In logistic regression, what is the role of the logistic function (sigmoid function)?

↳ it Convert distance into Probab.

~~A)~~ It transforms the independent variables.
B) It models the relationship between the dependent and independent variables.
~~C)~~ It converts the log-odds into probabilities. × $\log odd = \beta_0 + \beta_1 x \dots$
~~D)~~ It calculates the likelihood of the data.

Logistic Regression

Which term represents the natural logarithm of the odds of an event occurring in logistic regression? log(odds)

A) Odds ratio

B) Probability

C) Log-odds or logit

D) Coefficient

find β's.

RSS

$$\min \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

a) this is used to find β

b) this is used to access the fitness of model
goodness of model.

Logistic Regression

What is the likelihood function used for in logistic regression?

(a)

A) To estimate the coefficients of the model.
B) To calculate the odds ratio.
C) To find the best threshold for classification.
D) To assess the fit of the model by maximizing the likelihood of the observed outcomes.

Logistic Regression

1. What kind of algorithm is logistic regression?

a) Cost function minimization

b) Ranking

c) Regression

d) Classification

Logistic Regression

6. Probability of an event occurring is 0.9. What is odds ratio?

a) 0.9:1

$\frac{.9}{.1}$ $\frac{9}{1}$

b) 9:1 ✓

c) 1:9

d) 1:0.9

#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 β_0, β_1, β_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?

(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

Linear Classification

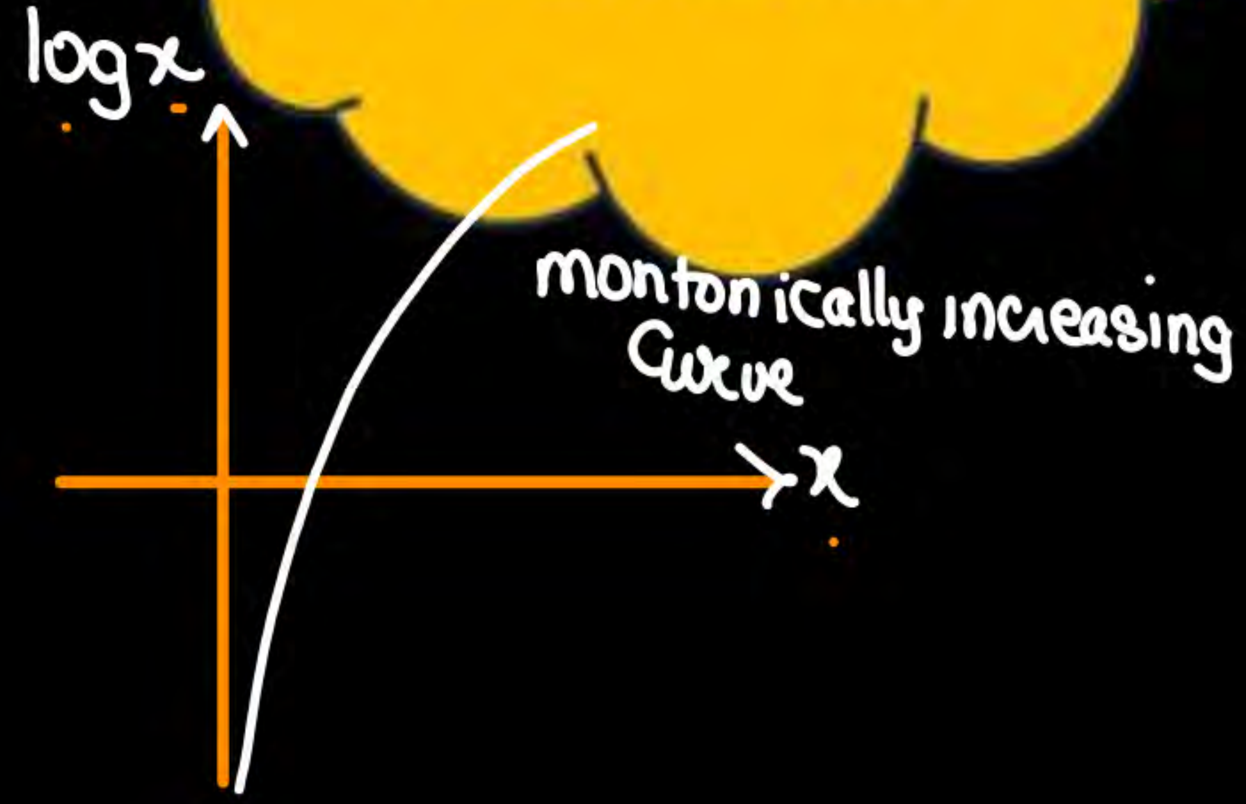
Logistic Regression

- **The Cost function**

Let $f(x)$ is any fxn $\Rightarrow$ • We have to maximize $f(x)$

max $f(x)$

How can we use log into this function



Logistic Regression

- **The Cost function**

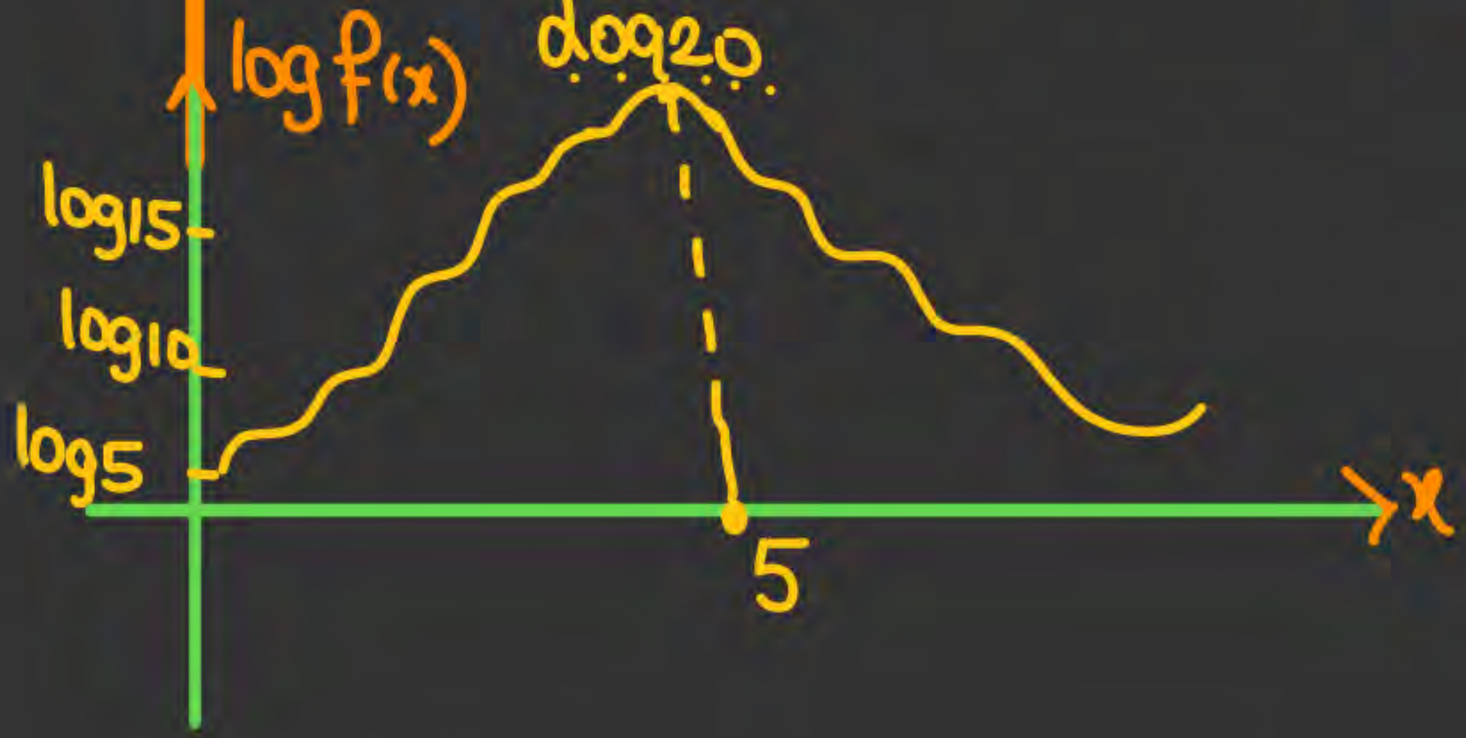
How can we use log into this function

max f(x) and max log f(x)

• log is a inc fxn

⇒ So value of x where f(x) is maxima

⇒ the log f(x) will also be maxima at same x.



log f(x)

- The pattern of plot of f(x) will be same as that log f(x)

Logistic Regression

- **The Cost function** How to find Best "β"

So $P(Y=1/X=x_i) = \dfrac{1}{1+e^{-x_i\beta}}$

It is the Probability that class of Point is '1' given the point is x_i

$P(Y=0/X=x_i) = \left(1 - \dfrac{1}{1+e^{-x_i\beta}}\right)$

It is the Probab that class of Point is '0' given that point is x_i

For any Point x_i we can get

$P(Y=1/X=x_i) \Rightarrow$ Probab that point belong to class 1

$$\frac{1}{1+e^{-\beta_1 x_i}}$$

$P(Y=0/X=x_i) = 1 - \frac{1}{1+e^{-\beta_1 x_i}}$

Probab that point belong to class 0

Logistic Regression

- **The Cost function** How to find Best "β"

$$\text{So } P(Y=1/X=x_i) = \frac{1}{1+e^{-x_i\beta}}$$

$$P(Y=0/X=x_i) = \left(1 - \frac{1}{1+e^{-x_i\beta}}\right)$$

- we have the training data, and here we know the class of all the points

Linear Classification

Logistic Regression

- **The Cost function**

So if the class of the point $y_p = 0$ then we maximise $\left(1 - \frac{1}{1+e^{-x_i\beta}}\right)$

" " " " " " $y_i = 1$ " " " $\frac{1}{1+e^{-x_i\beta}}$

Linear Classification

Logistic Regression

- **The Cost function**

5 data Point

$x_1 \rightarrow 1 \rightarrow \dfrac{1}{1+e^{-x_1\beta}}$

$x_2 \rightarrow 0 \rightarrow 1-\dfrac{1}{1+e^{-x_2\beta}}$

$x_3 \rightarrow 0 \rightarrow 1-\dfrac{1}{1+e^{-x_3\beta}}$

$x_4 \rightarrow 1 \rightarrow \dfrac{1}{1+e^{-x_4\beta}}$

$x_5 \rightarrow 1 \rightarrow \dfrac{1}{1+e^{-x_5\beta}}$

So algo

$$\max \left(\frac{1}{1+e^{-x_1\beta}} \left(1-\frac{1}{1+e^{-x_2\beta}}\right) \left(1-\frac{1}{1+e^{-x_3\beta}}\right) \cdot \frac{1}{1+e^{-x_4\beta}} \cdot \frac{1}{1+e^{-x_5\beta}} \right)$$

So we have to max product of Probab.

↳ if point is of class '1' → expression P_1

" " " " " '0' → " P_0

So if $y_i = 1 \Rightarrow \frac{1}{1+e^{-x_i\beta}}$

if $y_i = 0 \Rightarrow 1 - \frac{1}{1+e^{-x_i\beta}}$

The exp:-

$$\Rightarrow \max \prod_{i=1}^{N} \left(\frac{1}{1+e^{-x_i\beta}}\right)^{y_i} \left(1 - \frac{1}{1+e^{-x_i\beta}}\right)^{1-y_i}$$

if $y_i = 1$ i.e datapoint is of class 1

↳ this exp will have P_1

if $y_i = 0$ i.e data point belong to class 0

↳ this exp will have P_0.

Maximum Likelihood estimation

Cost fxn ⇒ $\cancel{\text{Max}}$. $\displaystyle\prod_{i=1}^{N} \left(\frac{1}{1+e^{-x_i\beta}}\right)^{y_i} \cdot \left(1-\frac{1}{1+e^{-x_i\beta}}\right)^{1-y_i}$

$\log a \cdot b \rightarrow \log a + \log b$

Same as

$$\rightarrow \text{Max} \log_e \left(\prod_{i=1}^{N} \left(\frac{1}{1+e^{-x_i\beta}}\right)^{y_i} \left(1-\frac{1}{1+e^{-x_i\beta}}\right)^{1-y_i} \right)$$

Same as $\max \sum_{i=1}^{N} \log_e \left[\left[\frac{1}{1+e^{-x_i\beta}}\right]^{y_i} \left[1 - \frac{1}{1+e^{-x_i\beta}}\right]^{1-y_i} \right]$

$$\sum_{i=1}^{N} \log_e \left(\frac{1}{1+e^{-x_i\beta}}\right)^{y_i} + \log_e \left(1 - \frac{1}{1+e^{-x_i\beta}}\right)^{1-y_i}$$

$$\max \sum_{i=1}^{N} y_i \log_e \left(\frac{1}{1+e^{-x_i\beta}}\right) + (1-y_i) \log_e \left(\frac{e^{-x_i\beta}}{1+e^{-x_i\beta}}\right)$$

$$\sum_{i=1}^{N} y_i \cancel{\log_e \frac{1}{1+e^{-x_i\beta}}} + (1-y_i)\log_e\left(e^{-x_i\beta}\right) + (1-y_i)\log\left(\frac{1}{1+e^{-x_i\beta}}\right)$$

• Max $\Rightarrow$ $$\boxed{\sum_{i=1}^{N}\left[\log_e \frac{1}{1+e^{-x_i\beta}} + (1-y_i)(-x_i\beta)\right]}$$

$$\log_e f(x) \downarrow \frac{1}{f(x)} \cdot f'(x)$$

$$L \Rightarrow \sum_{i=1}^{N}\left[y_i x_i\beta - x_i\beta - \log_e\left(1+e^{-x_i\beta}\right)\right]$$

$$\frac{\partial L}{\partial \beta} \Rightarrow \sum_{i=1}^{N}\left[y_i x_i - x_i + \frac{e^{-x_i\beta}}{1+e^{-x_i\beta}} x_i\right] = 0$$

So here will not get any expression of
B.

Logistic Regression

- **Extending the case for more than 2 classes... (not imp)**

Let us have 4 classes ⇒ let select any one class "4"

Parameter of 1st Sigmoid

$$\text{Now } \log_e \frac{P(Y=1/X=x)}{P(Y=4/X=x)} = x\beta_1$$

$$\log_e \frac{P(Y=2/X=x)}{P(Y=4/X=x)} = x\beta_2$$

$$\log_e \frac{P(Y=3|X=x)}{P(Y=4|X=x)} = x\beta_3$$

Logistic Regression

- Extending the case for more than 2 classes... (not imp)

$$\underbrace{P(Y=1|X=x)} + P(Y=2|X=x) + P(Y=3|X=x) + \underbrace{P(Y=4|X=x)} = 1$$

$$e^{x\beta_1} \cdot P_4 + e^{x\beta_2} \cdot P_4 + e^{x\beta_3} \cdot P_4 + P_4 = 1$$

$$P_4 = \frac{1}{1+\sum_{j=1}^{3} e^{x\beta_j}}, \quad P_1 = \frac{e^{x\beta_1}}{1+\sum_{j=1}^{3} e^{x\beta_j}}, \quad P_2 = \frac{e^{x\beta_2}}{1+\sum_{j=1}^{3} e^{x\beta_j}}$$

Final Result

Case of 4 class we have 3 sigmoids

$$P_4 = \frac{1}{1+\sum_{j=1}^{3} e^{x\beta_j}}$$

$$P_3 = \frac{e^{x\beta_3}}{1+\sum_{j=1}^{3} e^{x\beta_j}}$$

$$P_1 = \frac{e^{x\beta_1}}{1+\sum_{j=1}^{3} e^{x\beta_j}}$$

$$P_2 = \frac{e^{x\beta_2}}{1+\sum_{j=1}^{3} e^{x\beta_j}}$$

Conclusion ⇒ Simply for any new point 'x' find the probabilities P_1, P_2, P_3, P_4 ; Whichever is max, assign that class to the point.

Logistic Regression

- Extending the case for more than 2 classes... (not imp)

- $\underbrace{\text{If we have K classes}} \Rightarrow$ K-1 Sigmoids

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

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

Logistic Regression

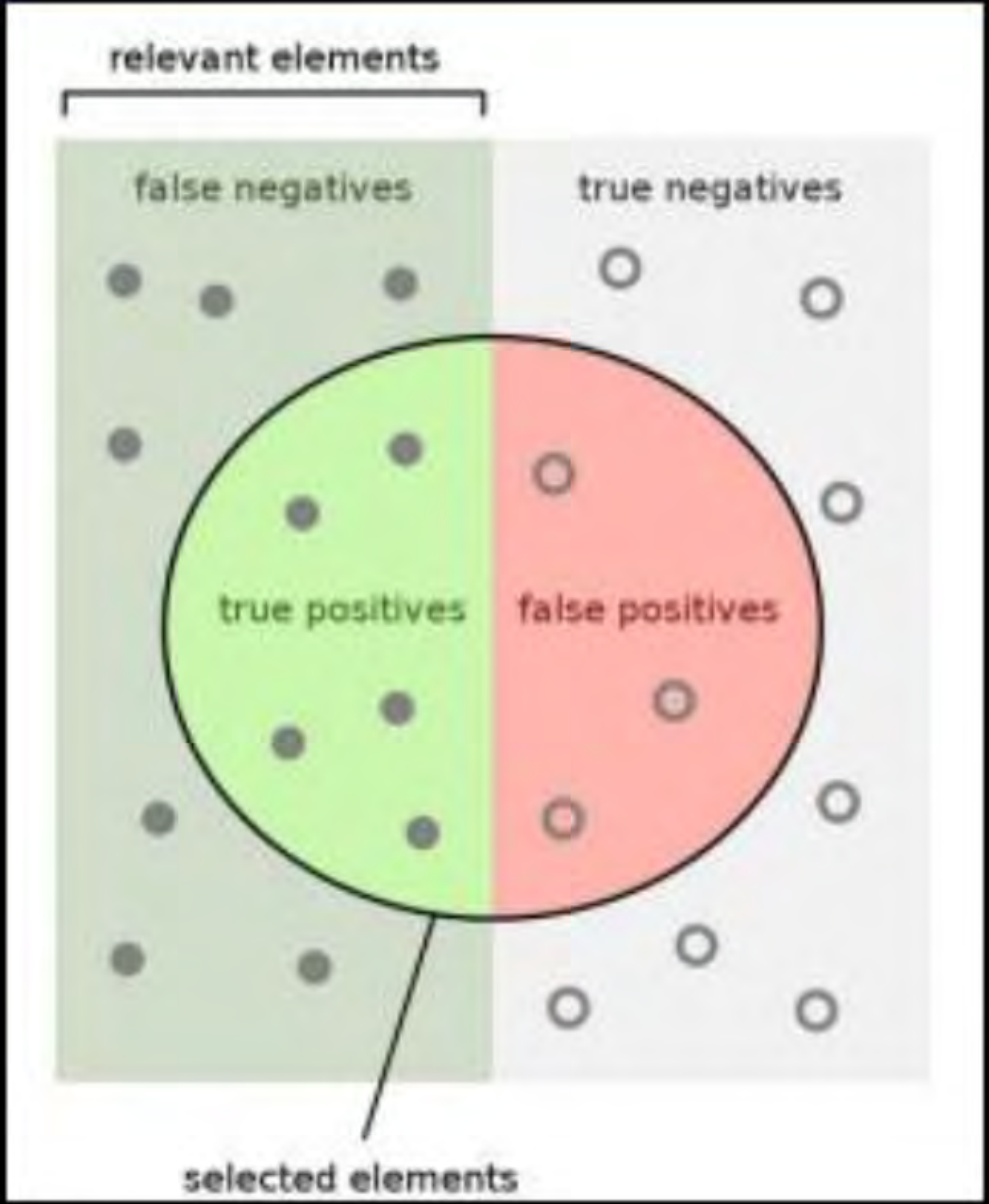
- **What is Confusion Matrix**

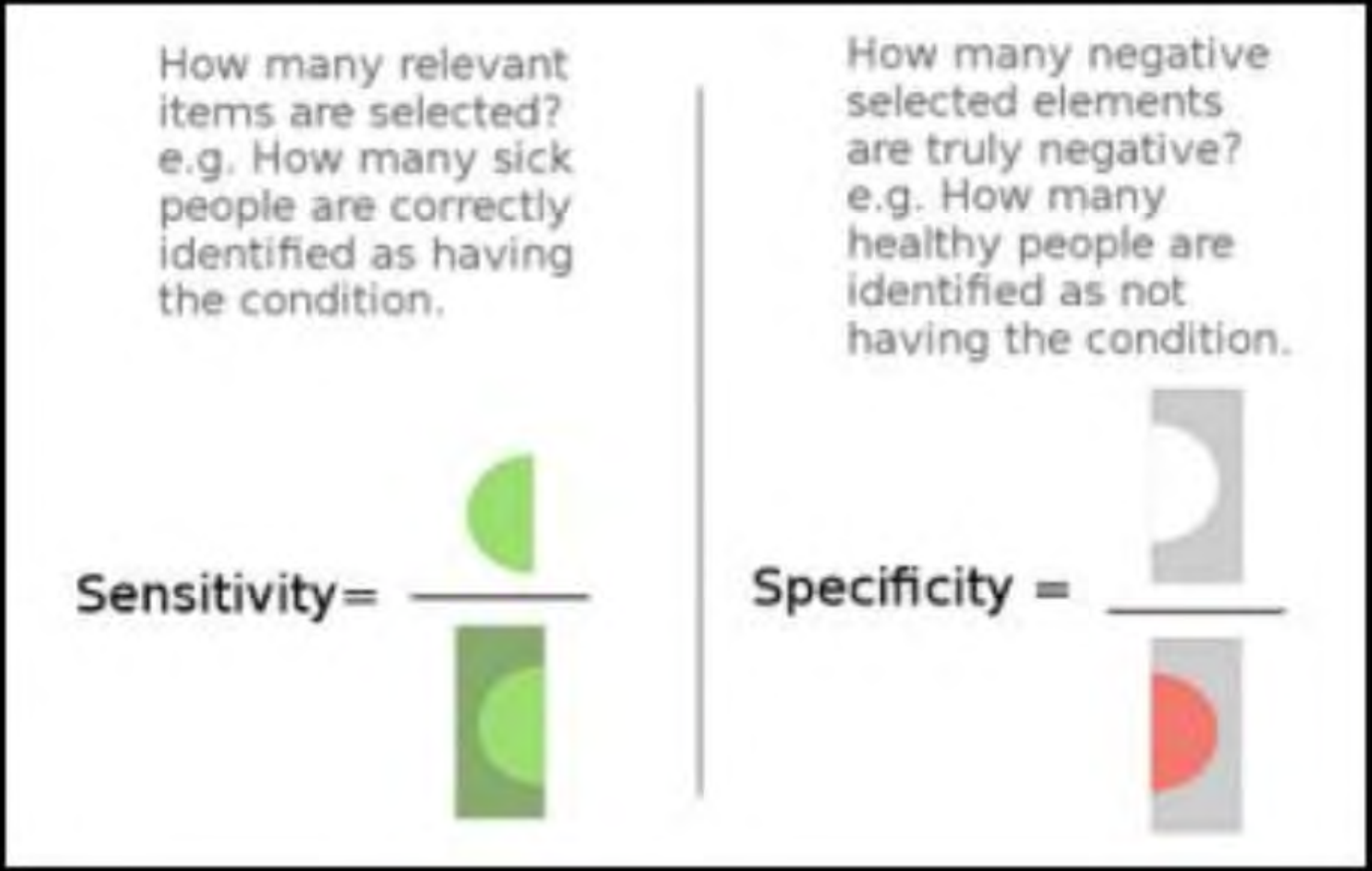
What is ROC curve (receiver operating characteristic curve)

- 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)





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:**

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

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

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:

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

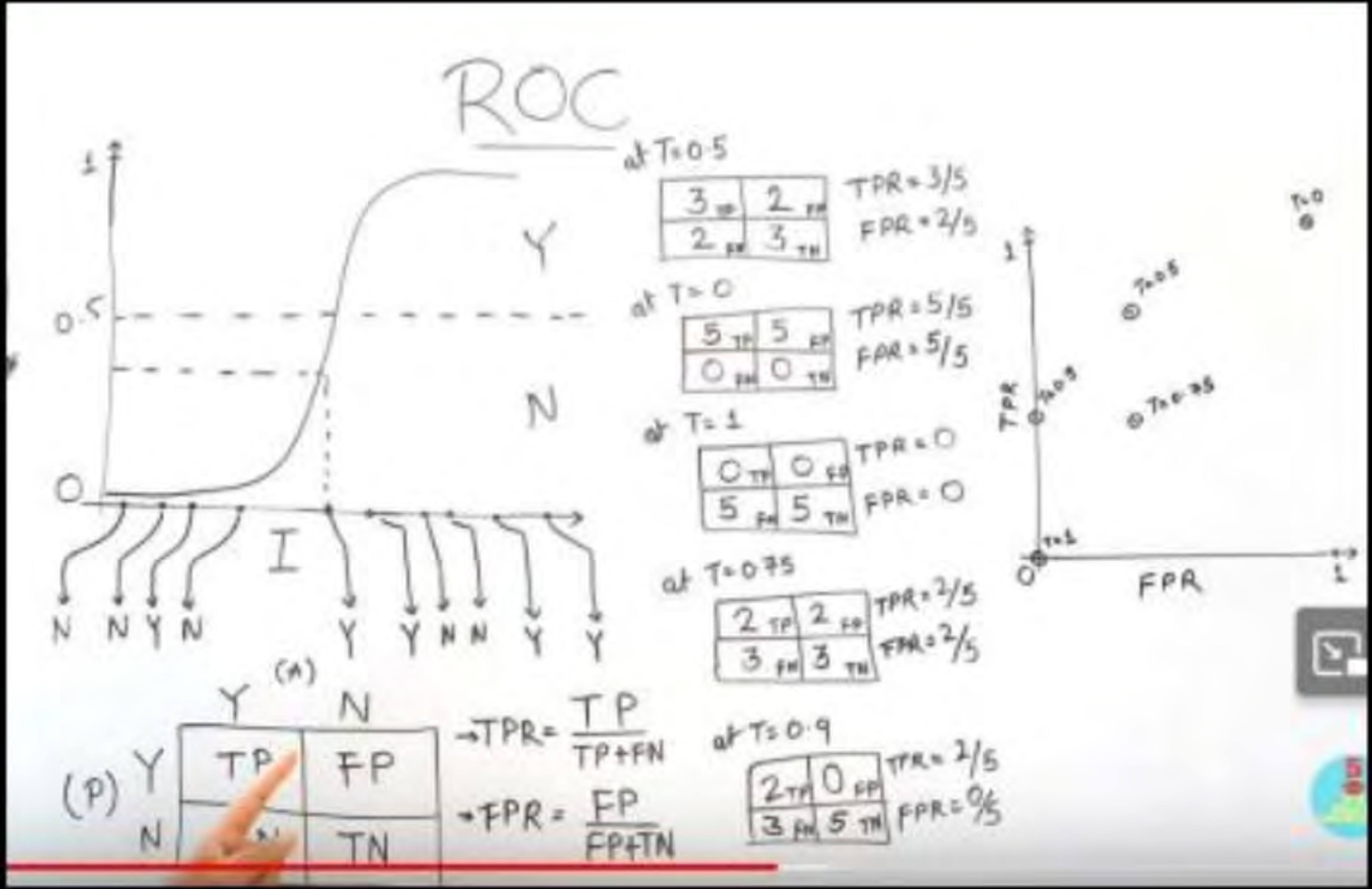
False Positive Rate (**FPR**) is defined as follows:

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

Linear Classification

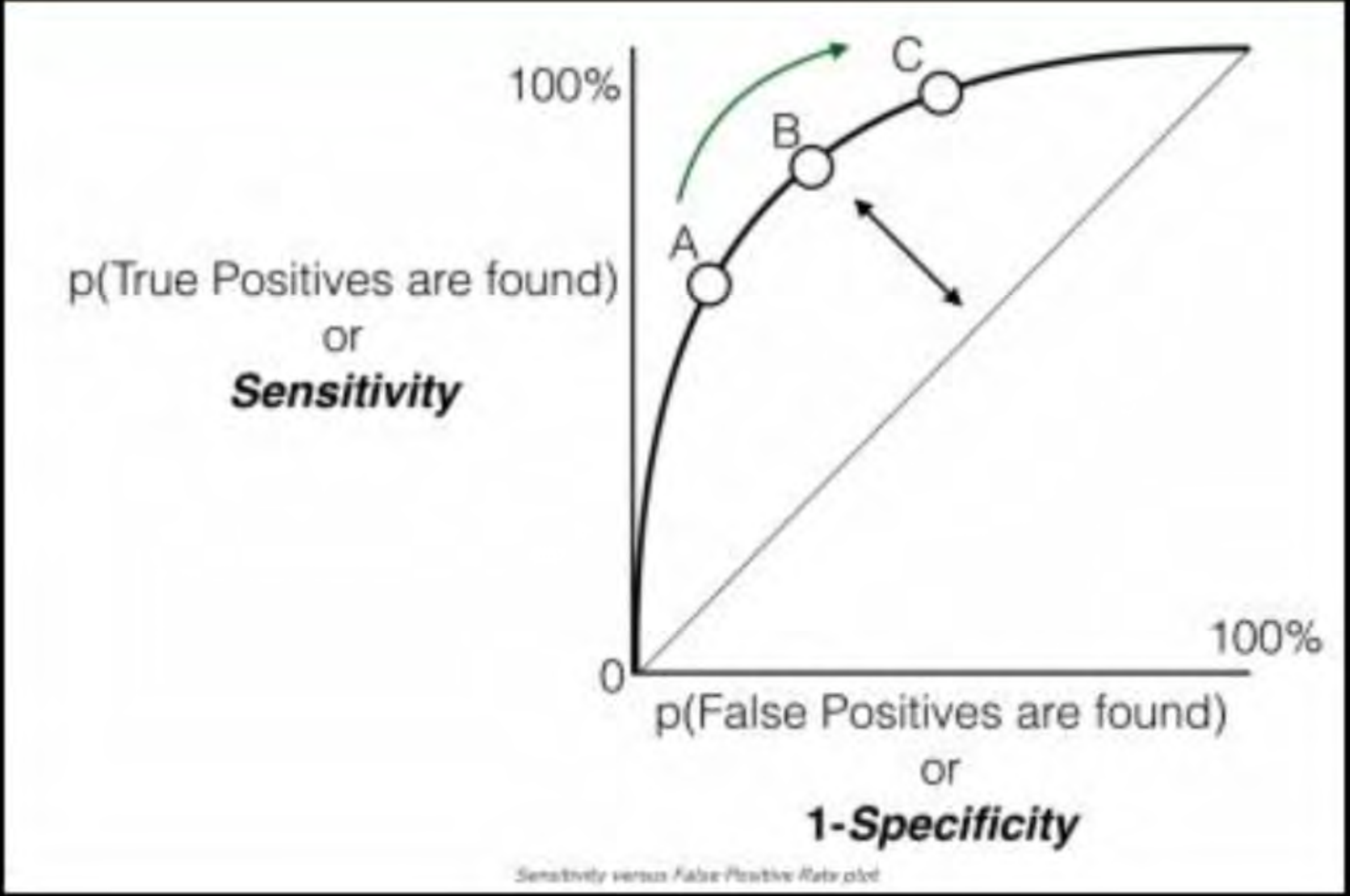
What is ROC curve (receiver operating characteristic curve) an example

The curve between TPR and FPR.



Linear Classification

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



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.

What is AUC (Area under the curve)

2 mins Summary

Topic

Topic

Topic

Topic

Topic

THANK - YOU