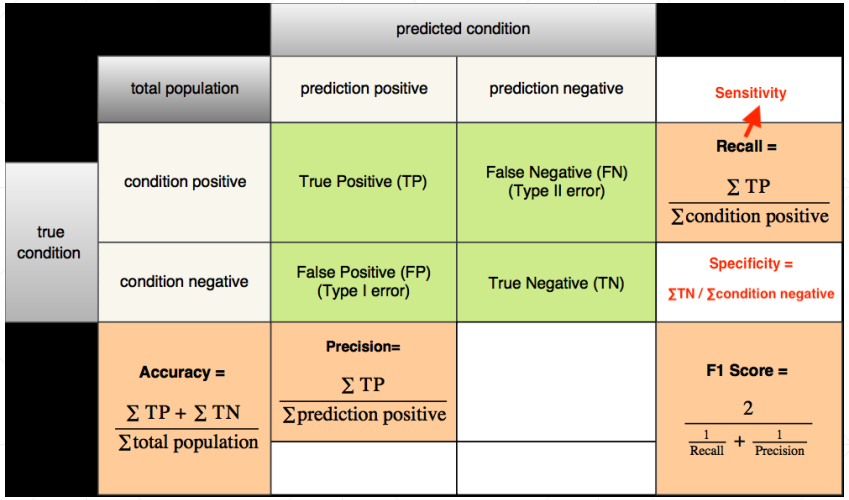
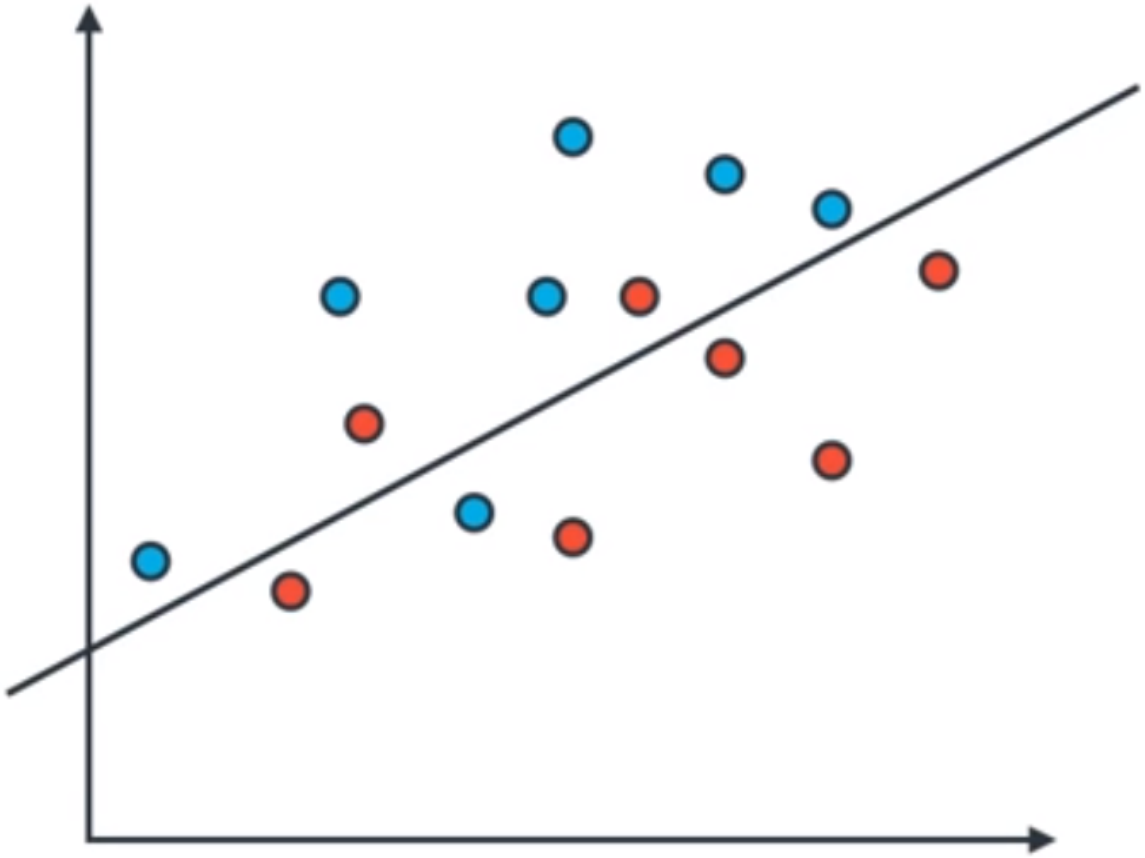
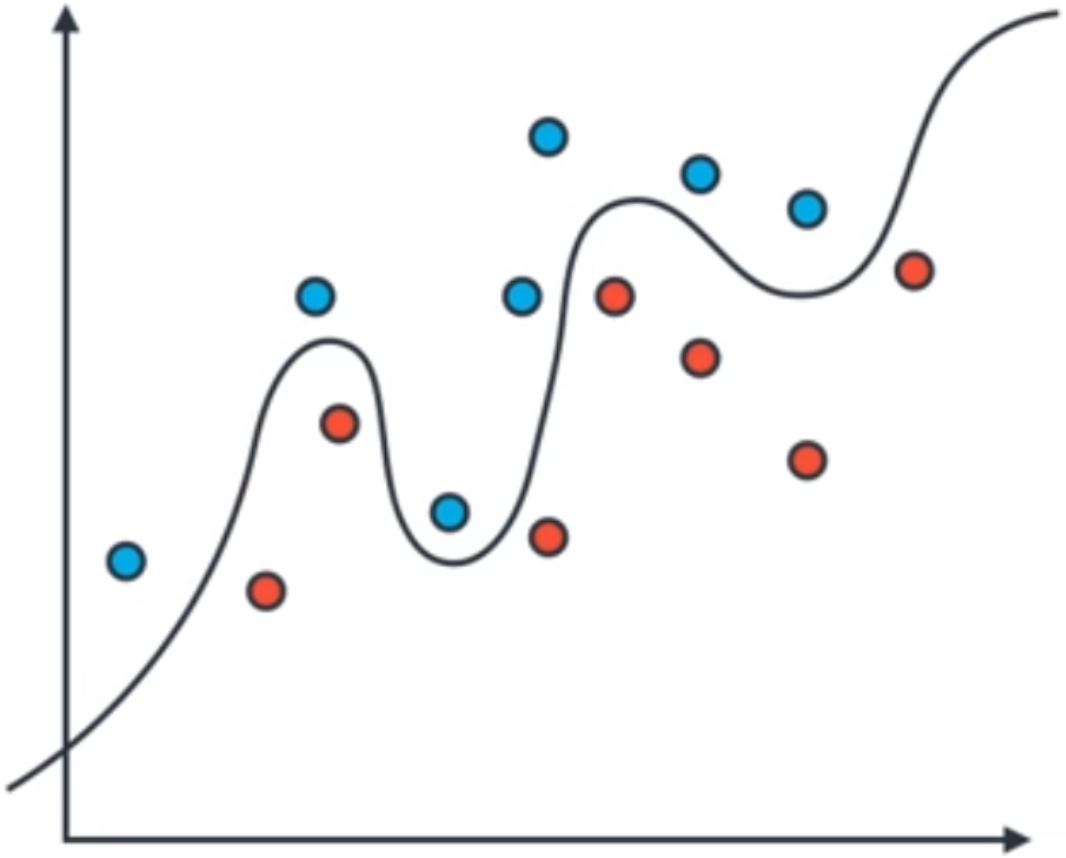


Evaluation metrics

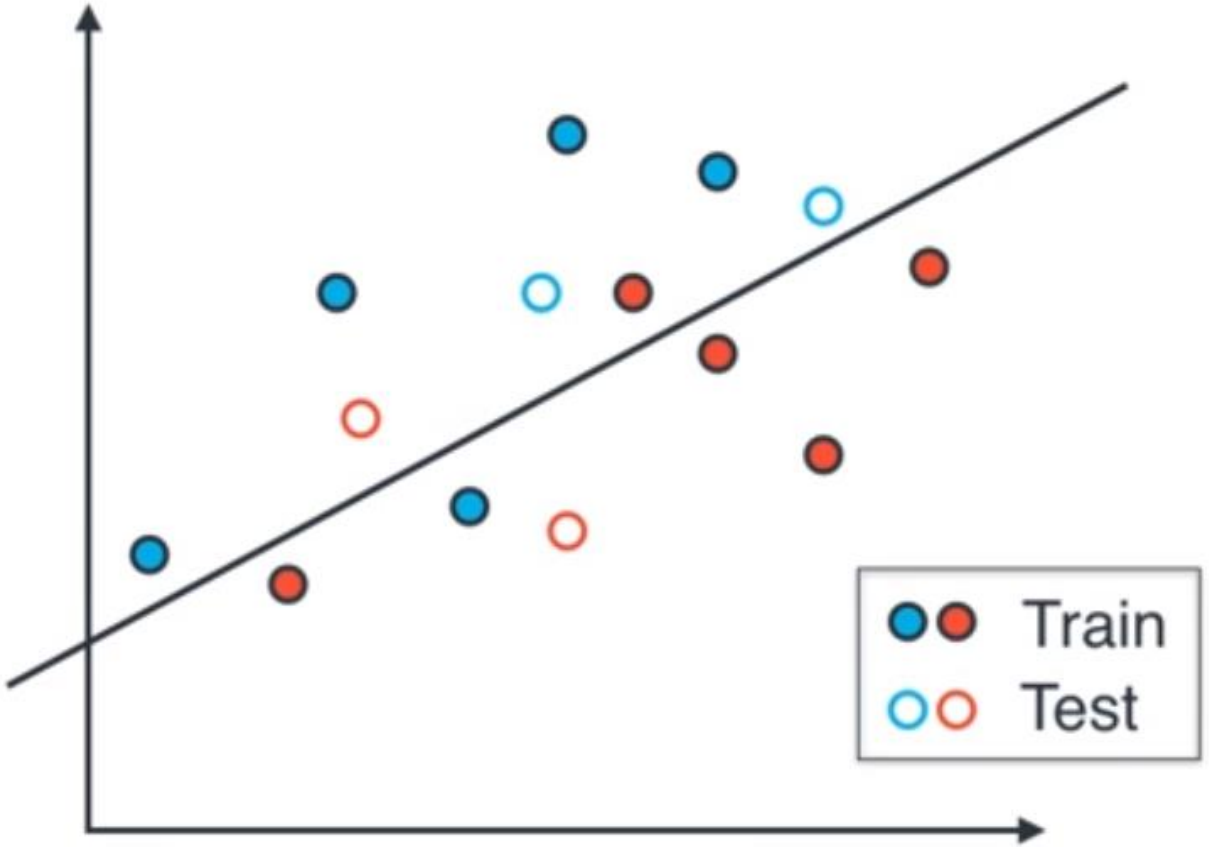
~Abhishek Kumar

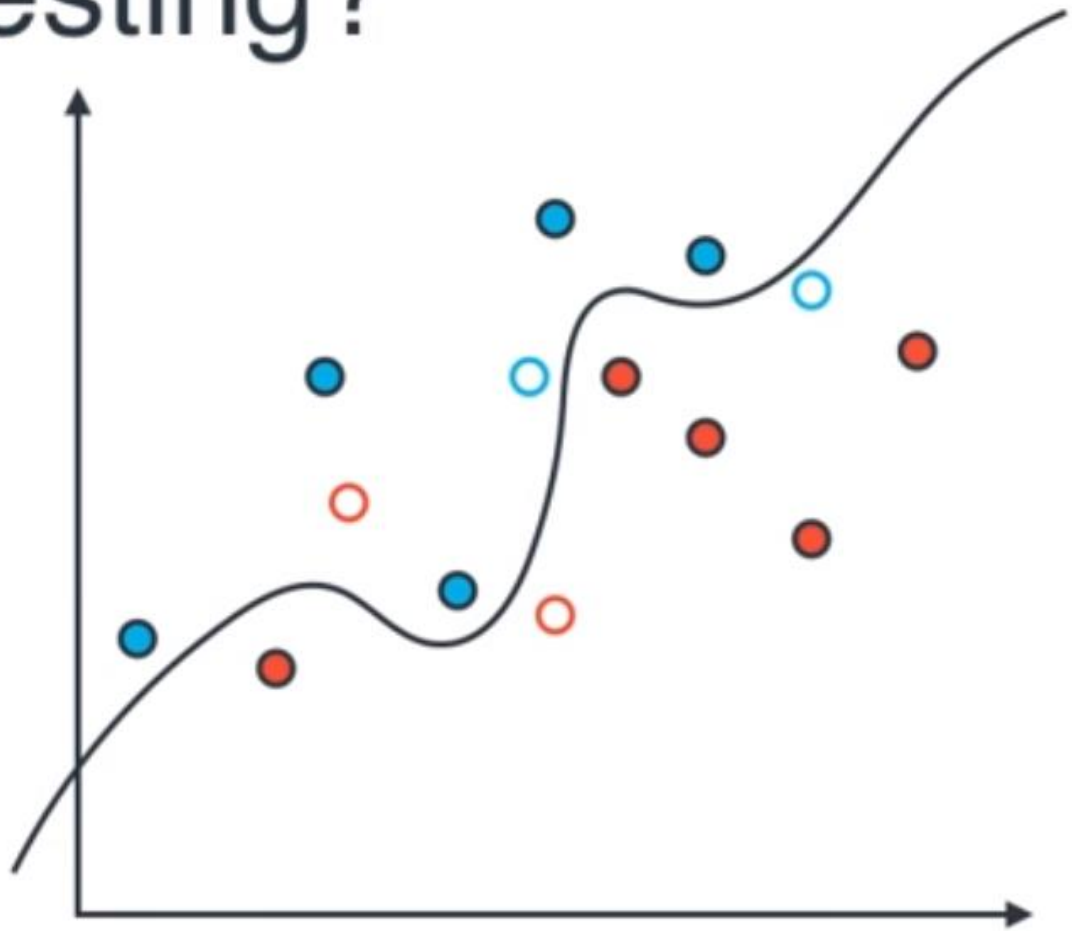
Which model is better



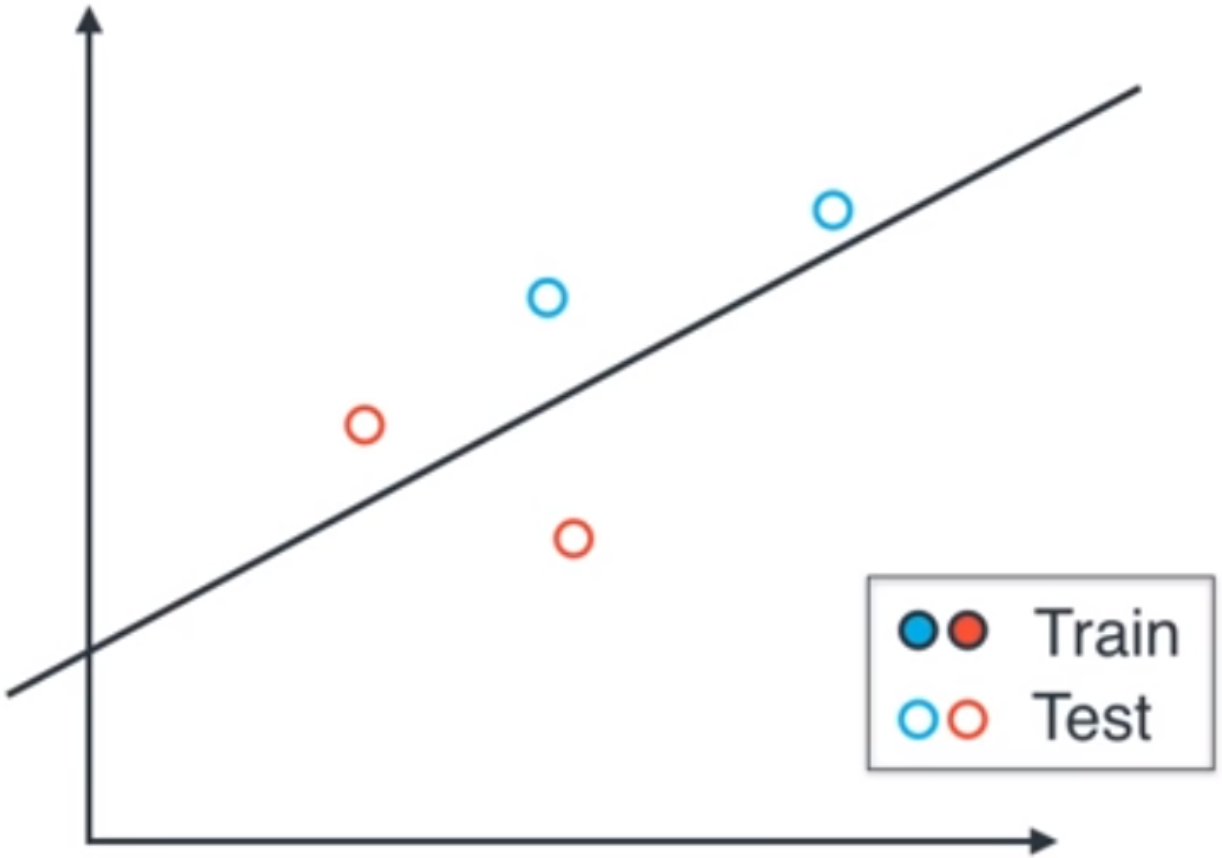


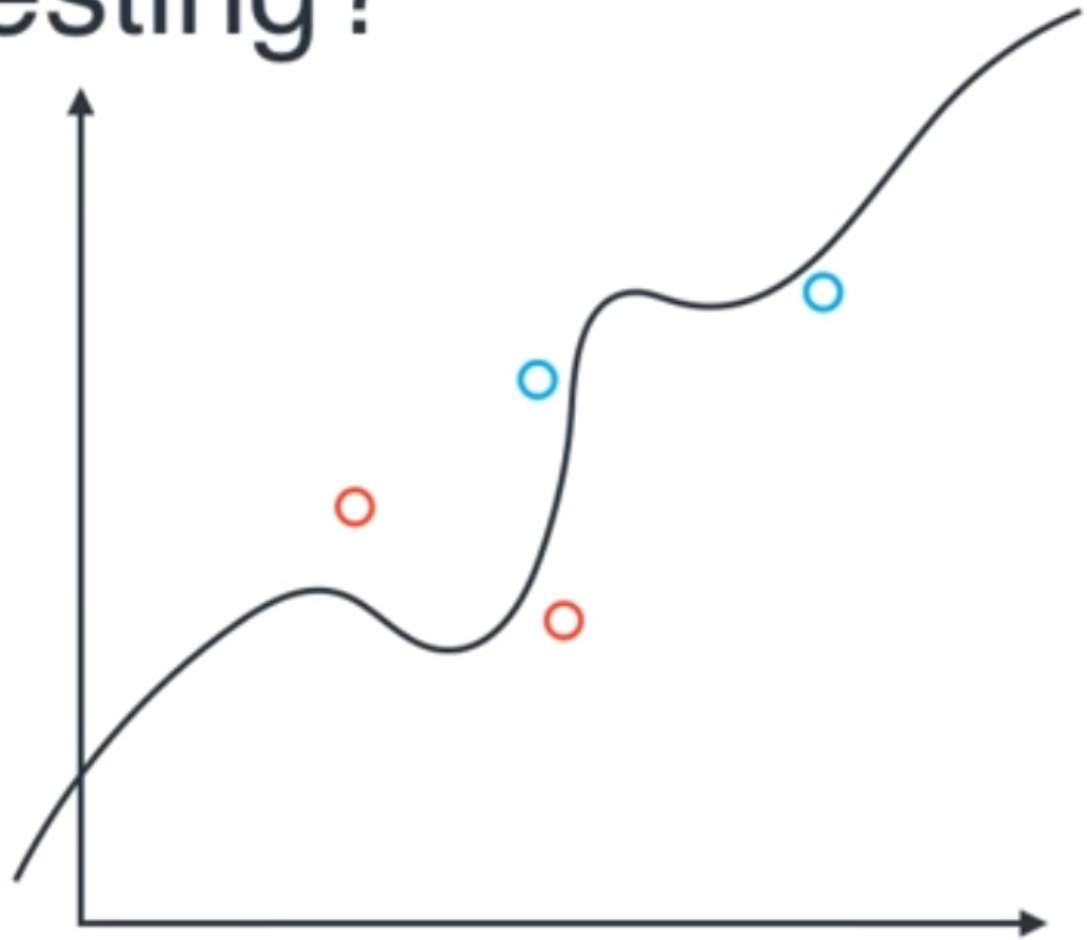
Why Testing?





Why Testing?

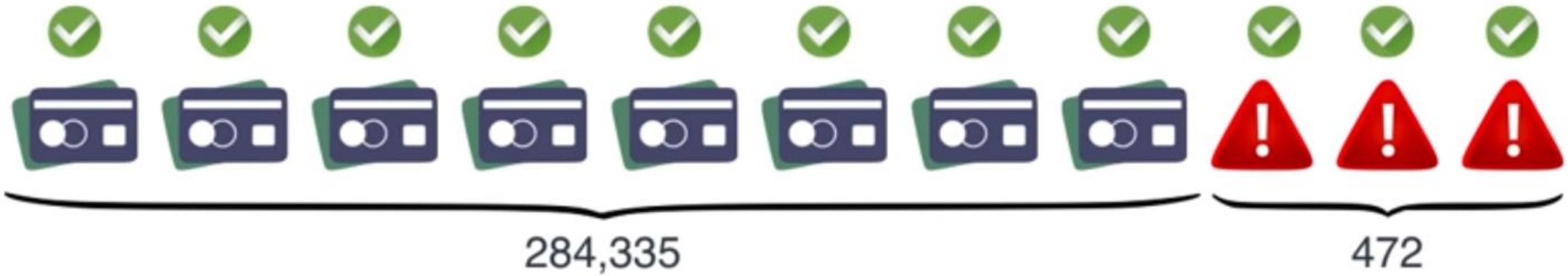





Training
Testing

K-Fold Cross Validation

Credit Card Fraud



Model: All transactions are good.

$$\text{Correct} = \frac{284{,}335}{284{,}807} = 99.83\%$$

	Diagnosed Sick	Diagnosed Healthy
Sick	True positive 	False Negative 
Healthy	False Positive 	True Negative 

	Diagnosed Sick	Diagnosed Healthy
Sick	True positive  	False Negative  
Healthy	False Positive  	True Negative  

Confusion Matrix

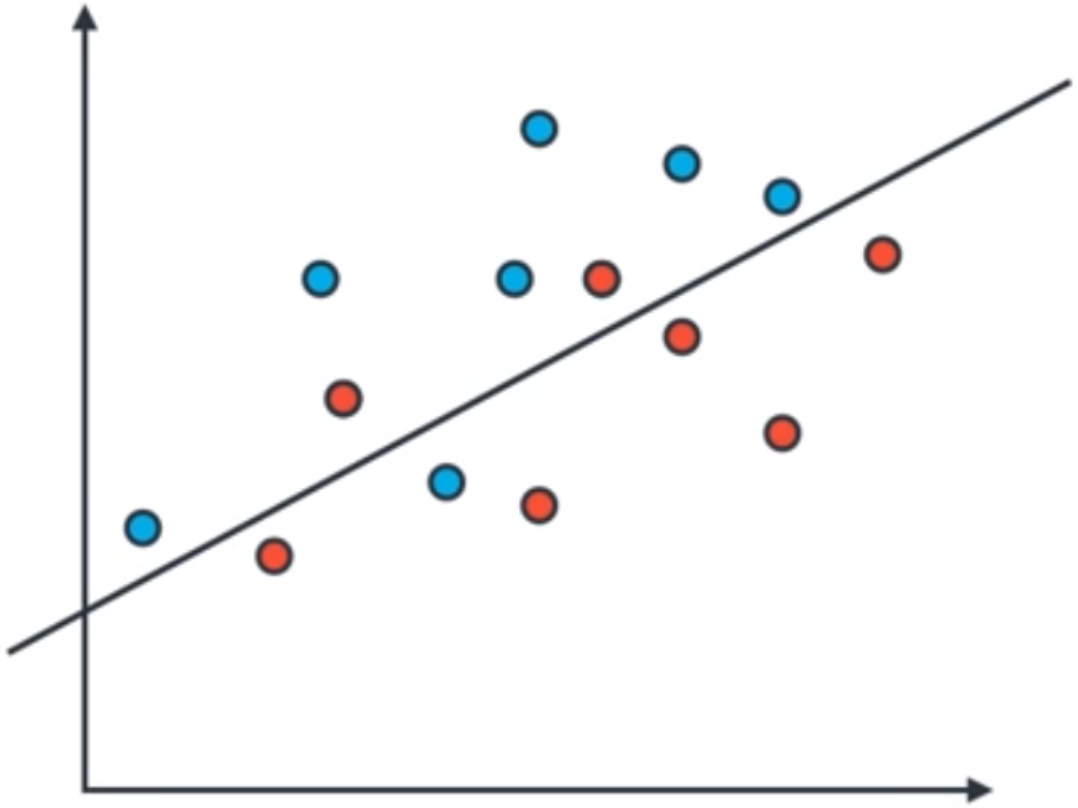
Diagnosis

Patients	Diagnosed sick	Diagnosed Healthy
Sick	1000 True positives	200 False Negatives
Healthy	800 False Positives	8000 True Negatives

Confusion Matrix

E-mail	Folder: Spam Folder	Folder: Inbox
Spam	100 True positives	170 False Negatives
Not spam	30 False Positives	700 True Negatives

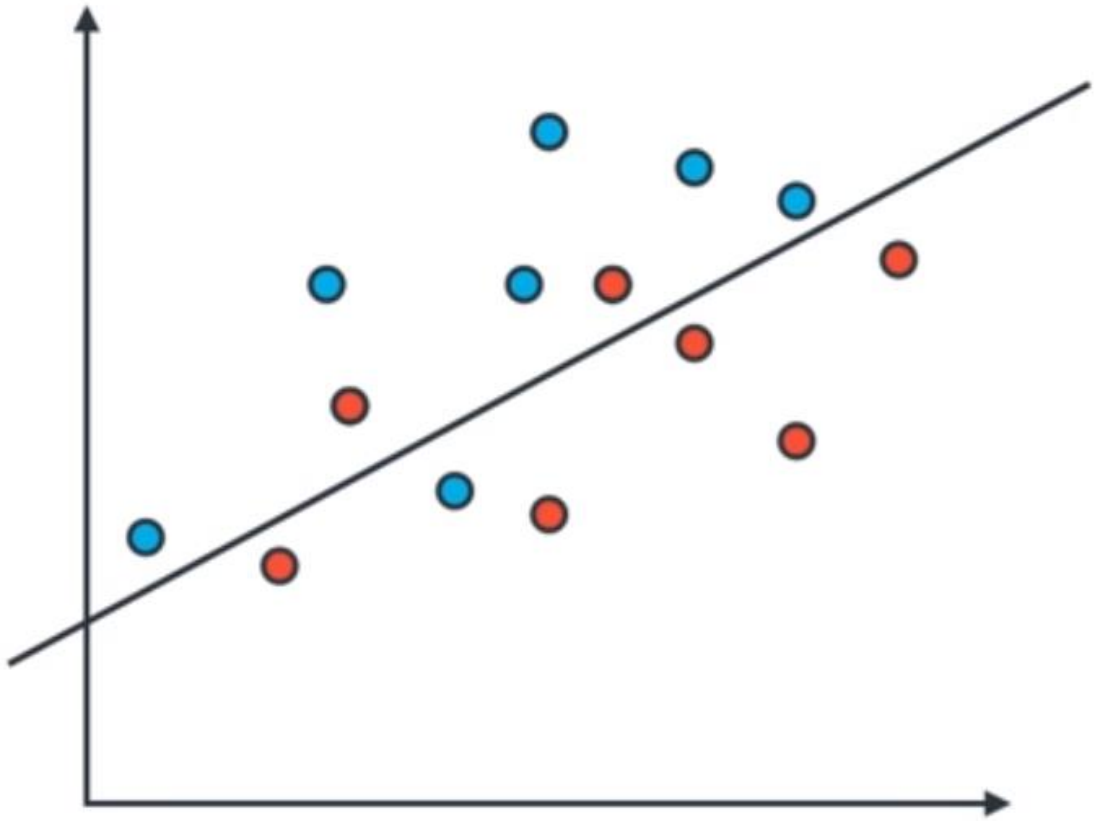
Confusion Matrix



Prediction

Data		Guessed Positive	Guessed Negative
	Positive		
	Negative		

Confusion Matrix



Prediction

Data	Guessed Positive	Guessed Negative
Positive	6 True positives	1 False Negatives
Negative	2 False Positives	5 True Negatives

Accuracy

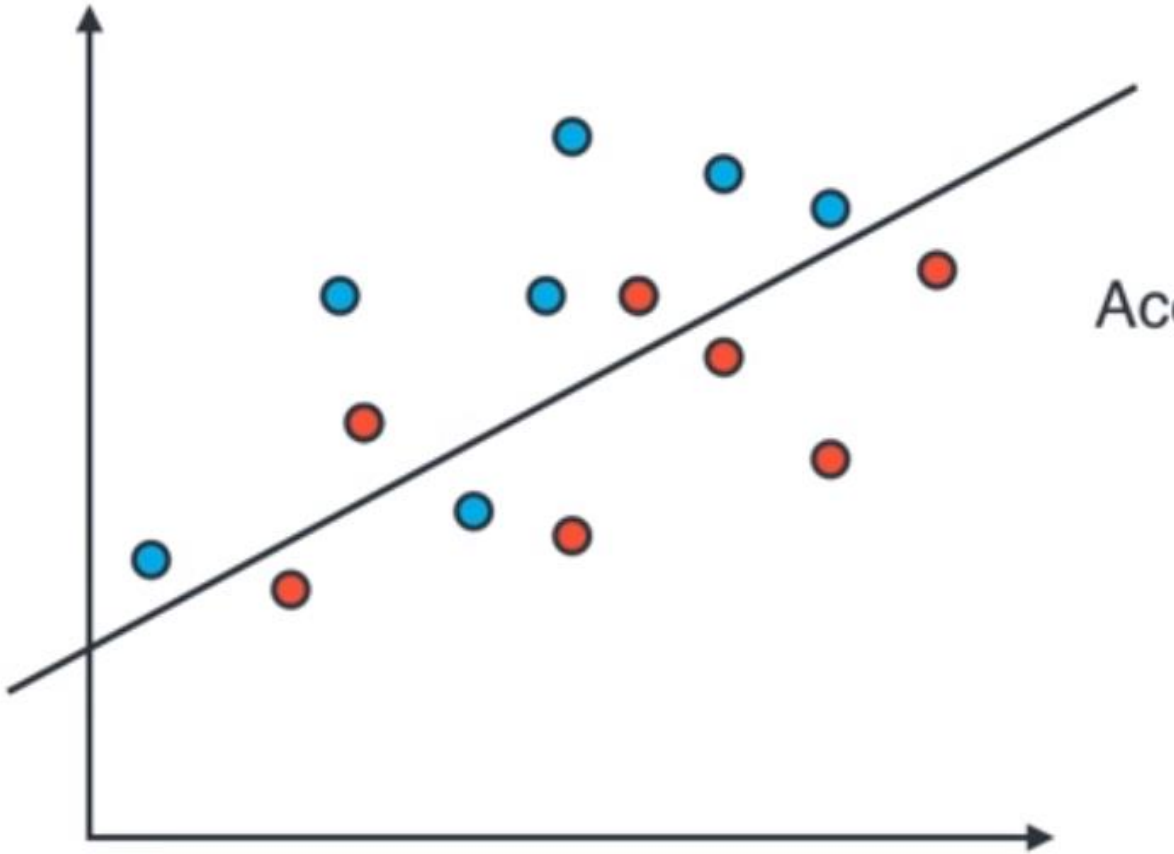
Diagnosis

Patients	Diagnosed sick	Diagnosed Healthy
Sick	1000	200
Healthy	800	8000

Accuracy: Out of the all the patients, how many did we classify correctly?

Accuracy =

Accuracy



Precision: Out of all the data, how many points did we classify correctly?

$$\text{Accuracy} = \frac{\text{Correctly Classified points}}{\text{All points}}$$

	Diagnosed Sick	Diagnosed Healthy
Sick		False Negative 
Healthy	False Positive 	

	Diagnosed Sick	Diagnosed Healthy
Sick		False Negative  
Healthy	False Positive  	

	Sent to Spam Folder	Sent to Inbox
Spam		False Negatives  
Not Spam	False Positives  	

	Sent to Spam Folder	Sent to Inbox
Spam		False Negatives  
Not Spam	False Positives 	

EVALUATION METRICS

Medical Model

False positives ok
False negatives **NOT** ok

Find all the sick people
Ok if not all are sick

High Recall

Spam Detector

False positives **NOT** ok
False negatives ok

You don't necessarily need to find all spam
But they better all be spam

High Precision

Precision

Patients \ Diagnosis	Diagnosed sick	Diagnosed Healthy
Sick	1000	200 ✗
Healthy	800	8000

Precision: Out of the patients we diagnosed with an illness, how many did we classify correctly?

$$\text{Precision} = \frac{1{,}000}{1{,}000 + 800} = 55.7\%$$

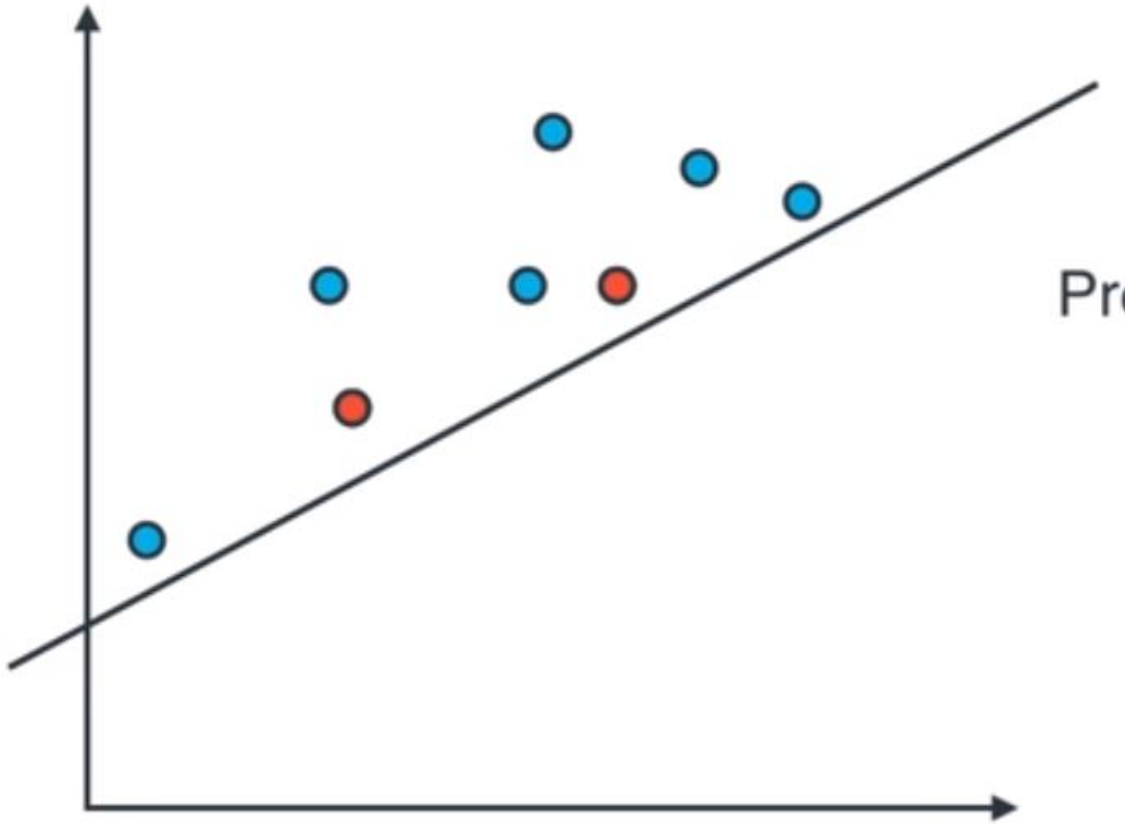
Precision

E-mail \ Folder	Spam Folder	Inbox
Spam	100	170
Not spam	30 ✗	700

Precision: Out of the all the e-mails, sent to the spam inbox, how many were actually spam?

$$\text{Precision} = \frac{100}{100 + 30} = 76.9\%$$

Precision



Precision: Out of the points we've predicted to be positive, how many are correct?

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False Positives}}$$

$$= \frac{\quad}{\quad}$$

Recall

Diagnosis

Patients	Diagnosed Sick	Diagnosed Healthy
Sick	1000	200 ✗
Is Healthy	800	8000

Recall: Out of the sick patients, how many did we correctly diagnose as sick?

$$\text{Recall} = \frac{1{,}000}{1{,}000 + 200} = 83.3\%$$

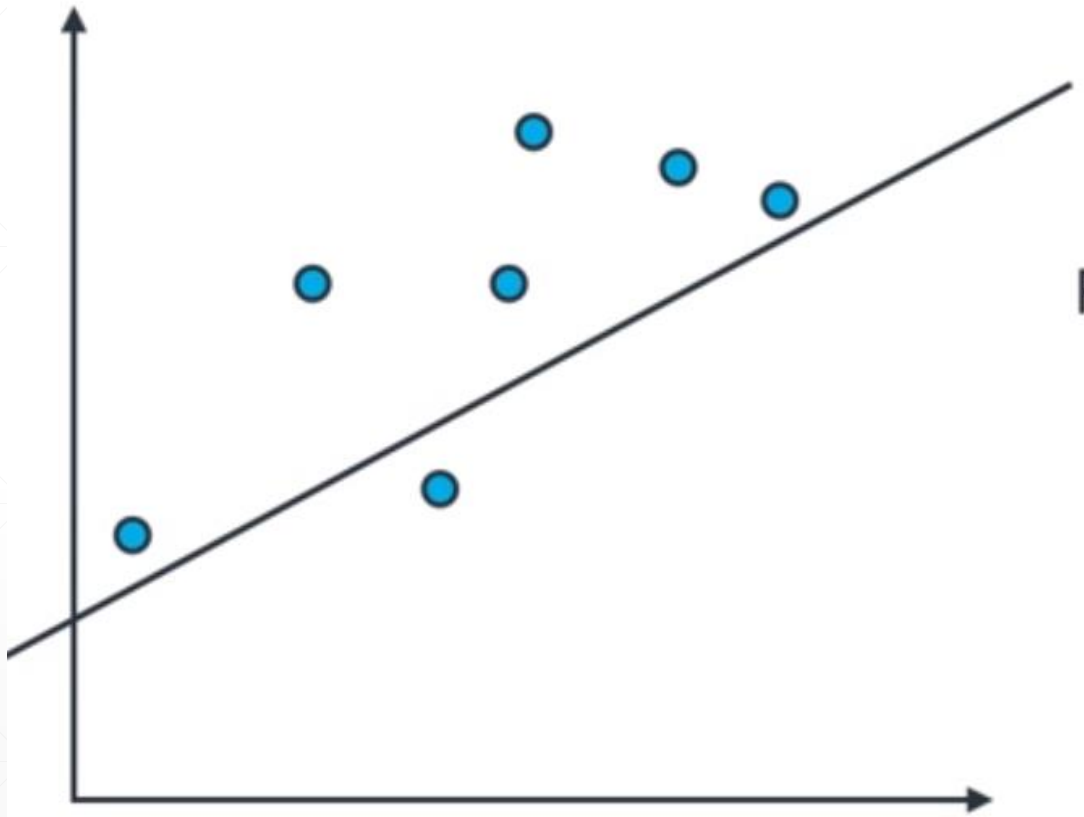
Recall

E-mail \ Folder	Spam Folder	Inbox
Spam	100	170
Not spam	30	700

Recall: Out of the all the spam e-mails, how many were correctly sent to the spam folder?

$$\text{Recall} = \frac{100}{100 + 170} = 37\%$$

Recall



Recall: Out of the points labelled positive, how many did we correctly predict?

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False Negatives}}$$

=

Precision and Recall

Medical Model

Precision: 55.7%
Recall: 83.3%

Spam Detector

Precision: 76.9%
Recall: 37%

F1 Score

Medical Model

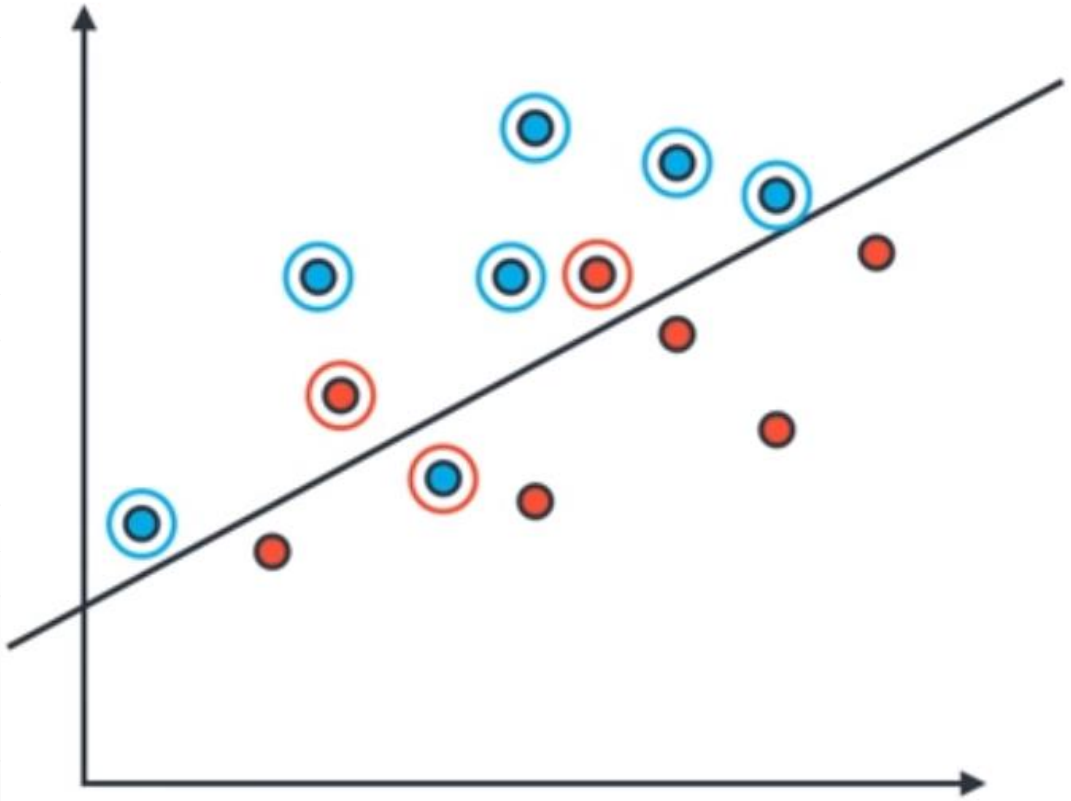
Precision = 55.7%

Recall = 83.3%

Average = 69.5%

$$\text{F1 Score} = \frac{2 \times 55.7 \times 83.3}{55.7 + 83.3} = 66.76\%$$

F1 Score



Precision = 75%

Recall = 85.7%

Average = 80.35

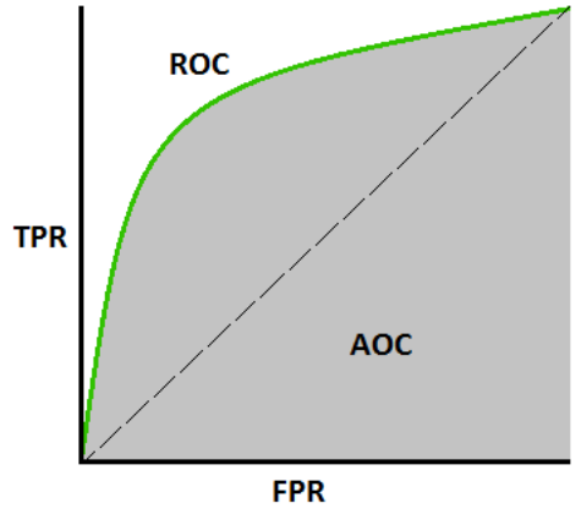
$$\text{F1 Score} = \frac{2 \times 75 \times 85.7}{75 + 85.7} = 80\%$$

		predicted condition		
	total population	prediction positive	prediction negative	Sensitivity
true condition	condition positive	True Positive (TP)	False Negative (FN) (Type II error)	**Recall =** $\frac{\sum TP}{\sum \text{condition positive}}$
	condition negative	False Positive (FP) (Type I error)	True Negative (TN)	Specificity = $\sum TN / \sum \text{condition negative}$
	Accuracy = $\frac{\sum TP + \sum TN}{\sum \text{total population}}$	**Precision=** $\frac{\sum TP}{\sum \text{prediction positive}}$		**F1 Score =** $\frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}$

Metric	Formula
True positive rate, recall	$\frac{TP}{TP+FN}$
False positive rate	$\frac{FP}{FP+TN}$
Precision	$\frac{TP}{TP+FP}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
F-measure	$\frac{2 \cdot precision \cdot recall}{precision + recall}$

AUC - ROC Curve

- ROC is a probability curve and AUC represents degree or measure of separability.
- It tells how much model is capable of distinguishing between classes.
- Higher the AUC, better the model
- Values between 0 to 1



TPR (True Positive Rate) / Recall /Sensitivity

$$\text{TPR /Recall / Sensitivity} = \frac{TP}{TP + FN}$$

Image 3

Specificity

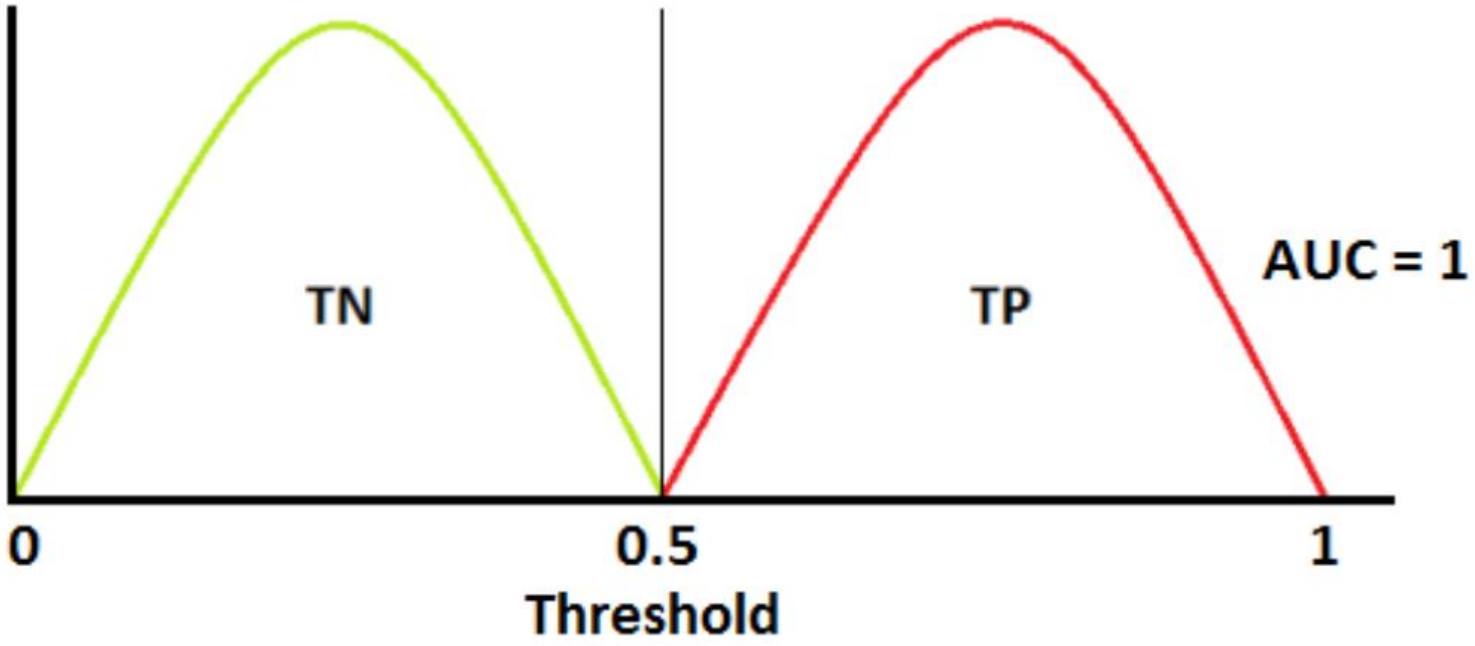
$$\text{Specificity} = \frac{TN}{TN + FP}$$

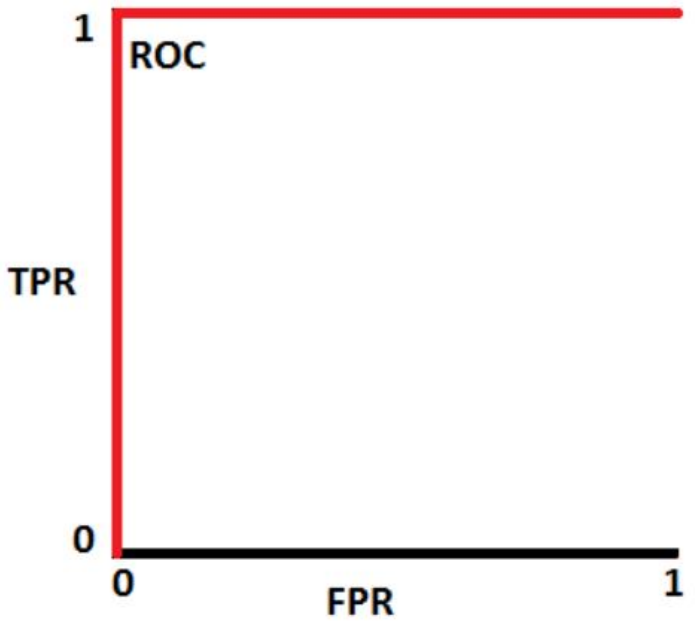
Image 4

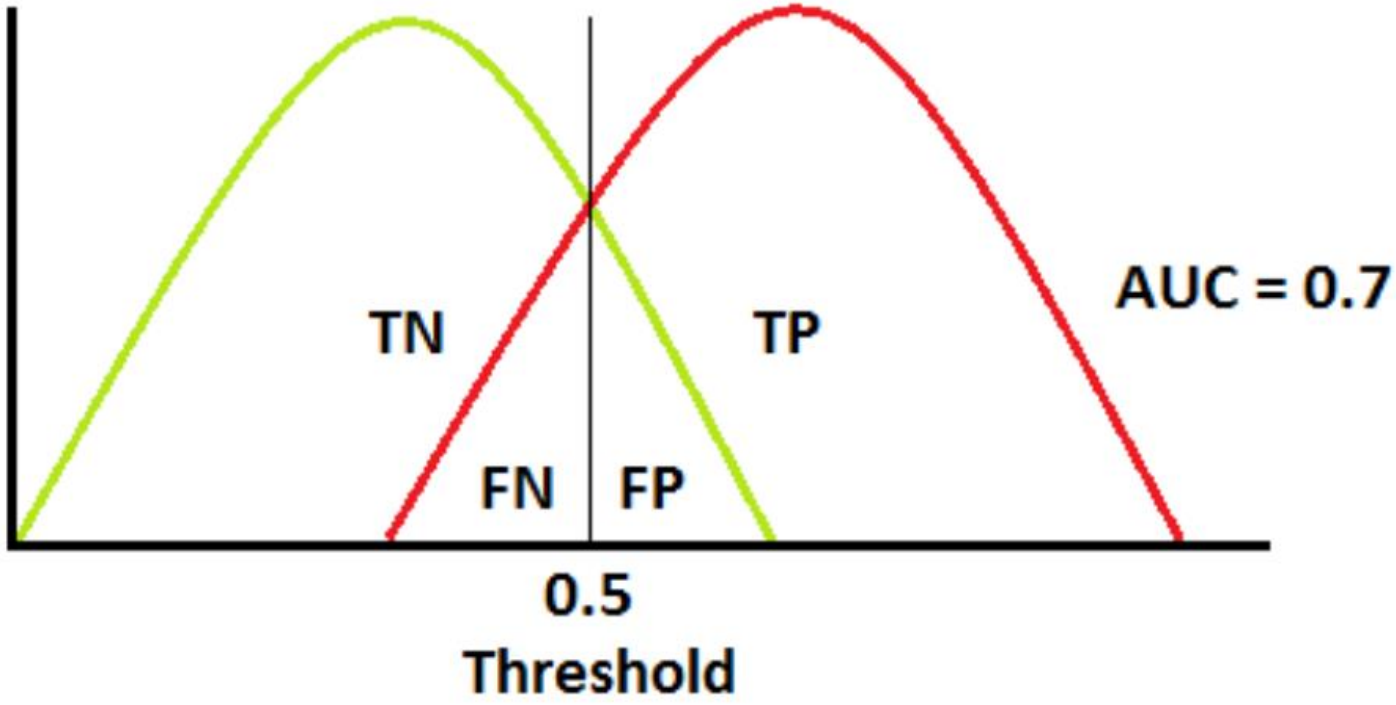
FPR

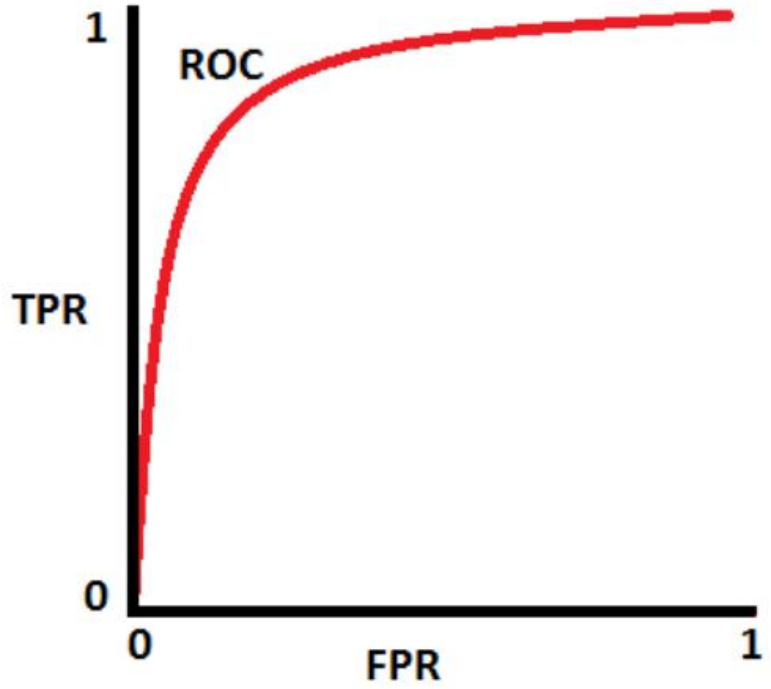
$$\text{FPR} = 1 - \text{Specificity}$$

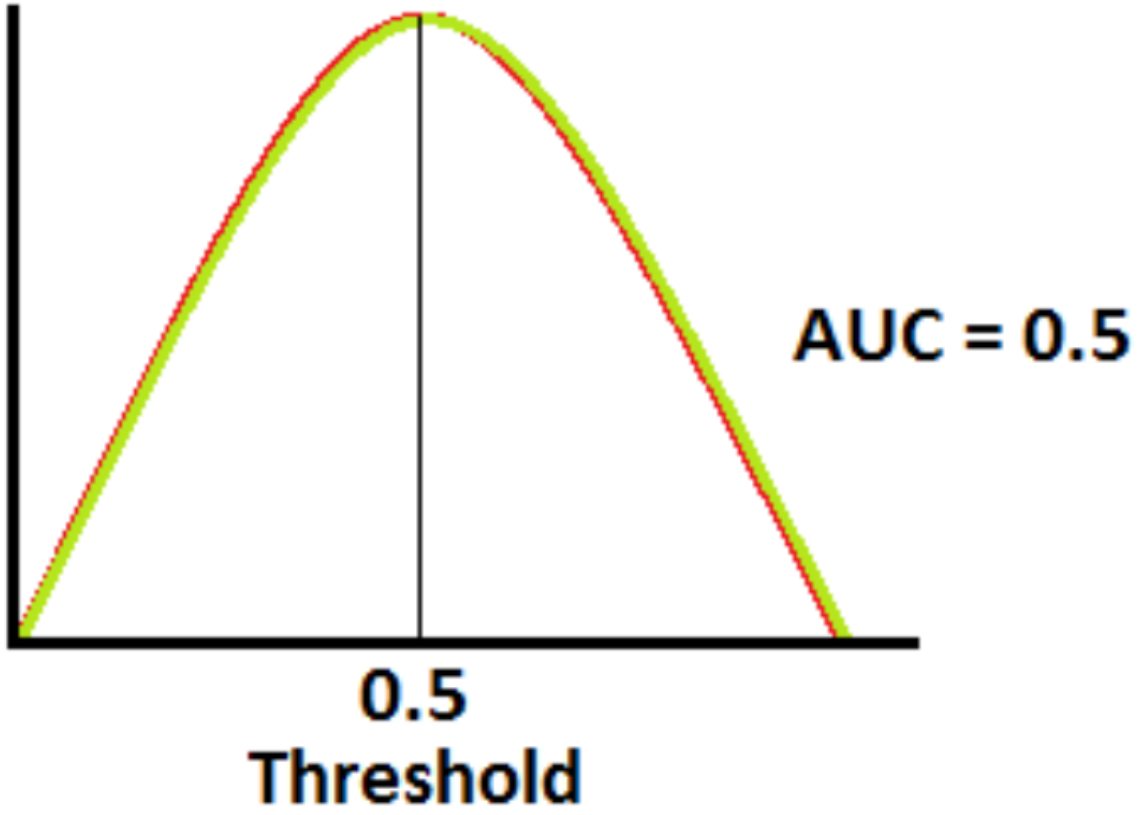
$$= \frac{FP}{TN + FP}$$

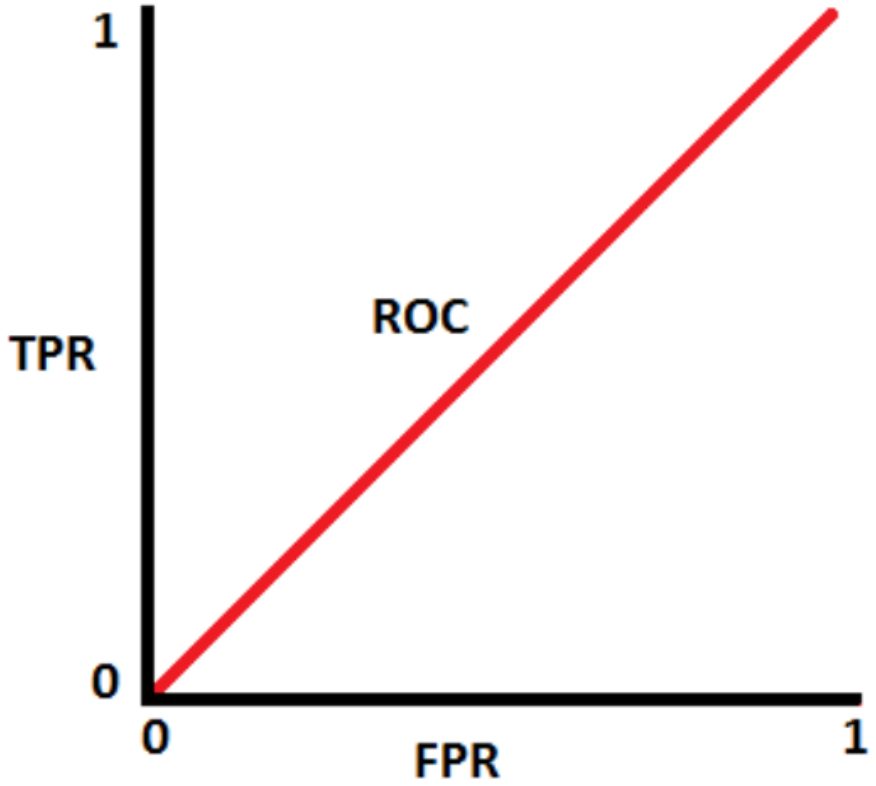

TN
TP
AUC = 1
0
0.5
1
Threshold

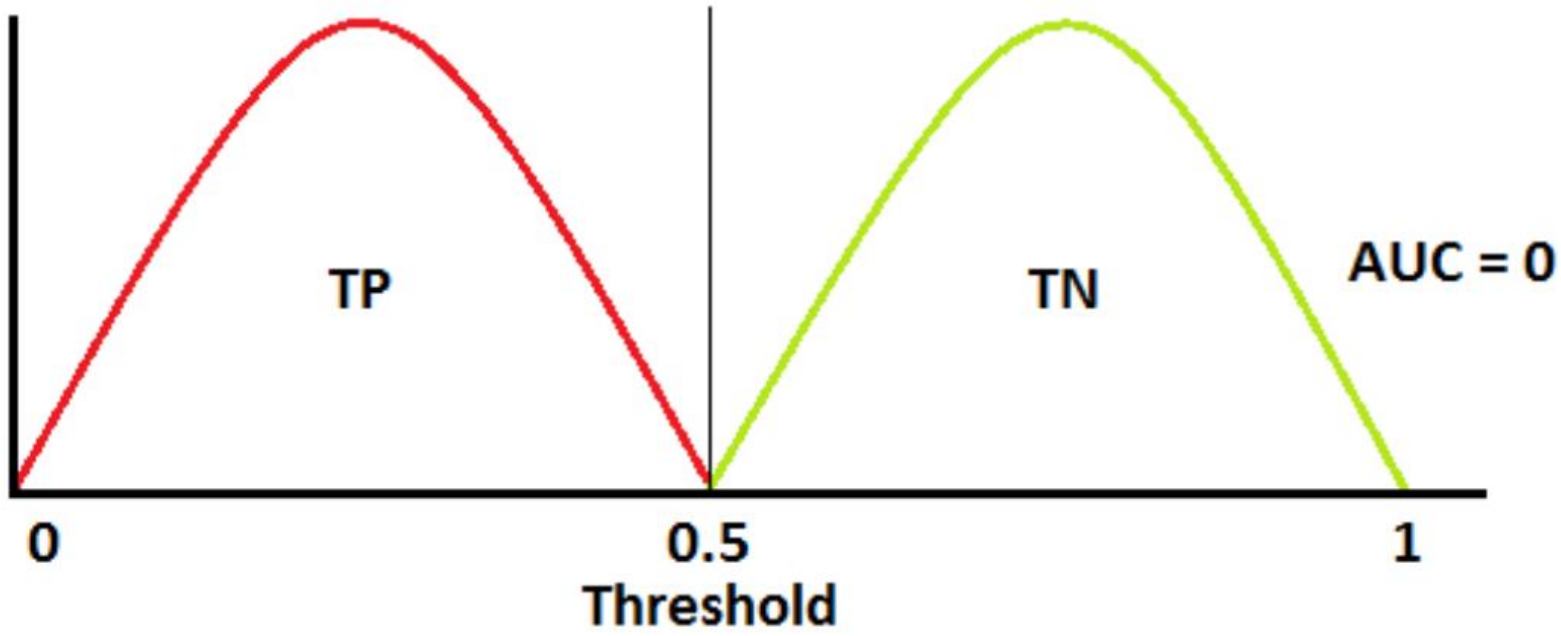

1
ROC
TPR
0
0
FPR
1

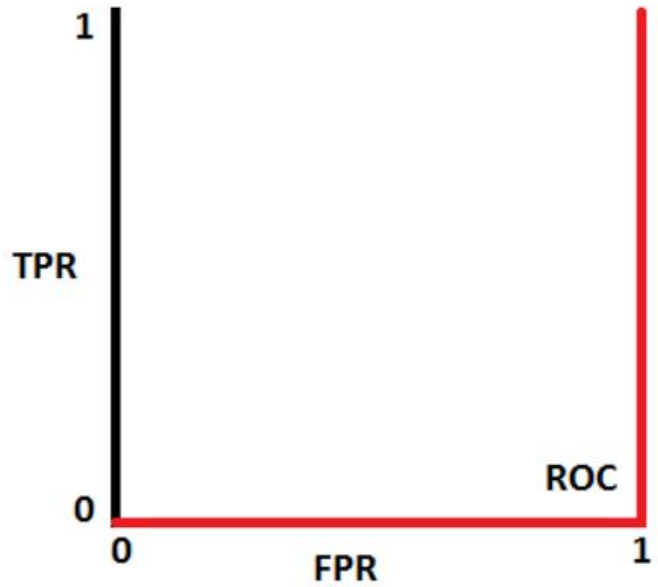

TN
TP
FN
FP
AUC = 0.7
0.5
Threshold


1
ROC
TPR
0
0
1
FPR

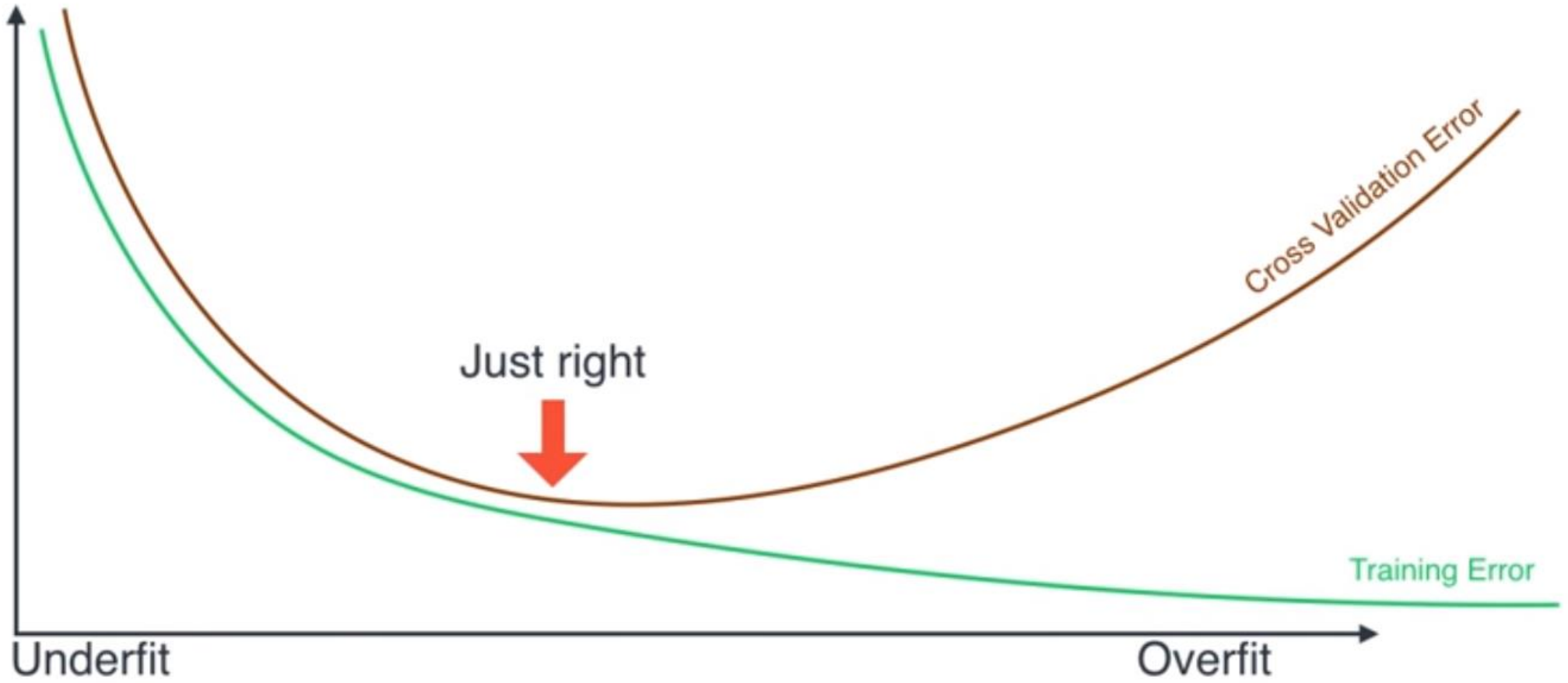

AUC = 0.5
0.5
Threshold


1
TPR
ROC
0
0
FPR
1


TP
TN
AUC = 0
0
0.5
Threshold
1


1
TPR
0
ROC
0
FPR
1

Model Complexity Graph



Parameters and Hyperparameters

Algorithm	Parameters	Hyperparameters
Random Forest	Features Thresholds	Number of trees Depth
Logistic Regression	Coefficients of the polynomial	Degree of the polynomial
Support Vector Machines	Coefficients	Kernel Gamma C
Neural Networks	Coefficients	Number of layers Size of layers Activation function

Discussion

Thank you!