Measuring Machine Learning Models

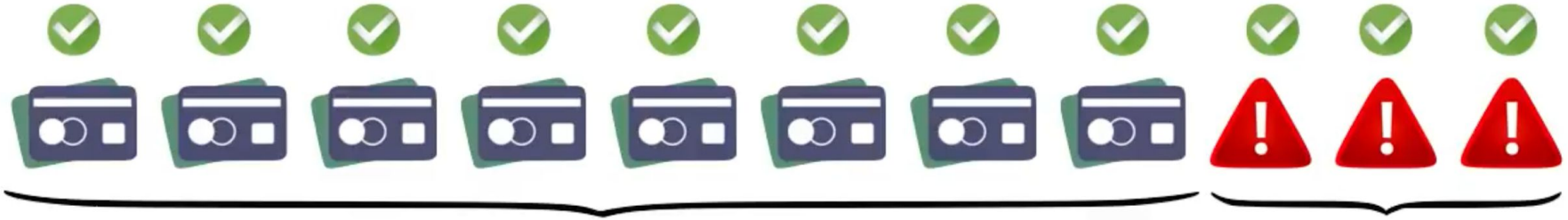
Prepared By: Dr.Mydhili K Nair, Professor, ISE Dept, RIT

For: Machine Learning Elective Class

Target Audience: Sem 6 Students

Term: Feb to June 2019

Credit Card Fraud

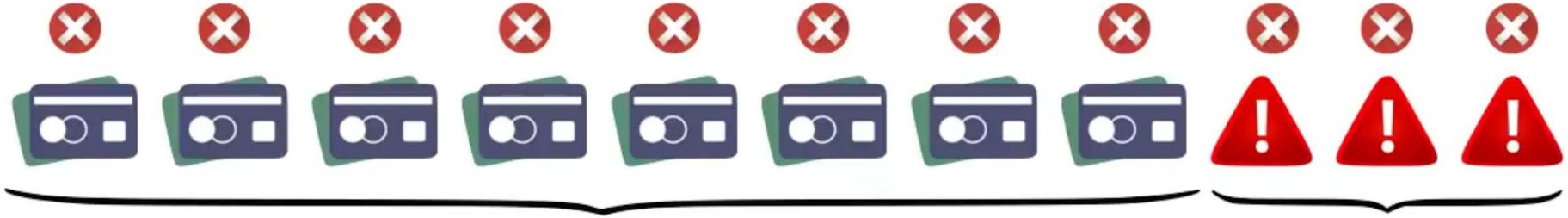


Model: All transactions are good.

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

Problem: I'm not catching any of the bad ones!

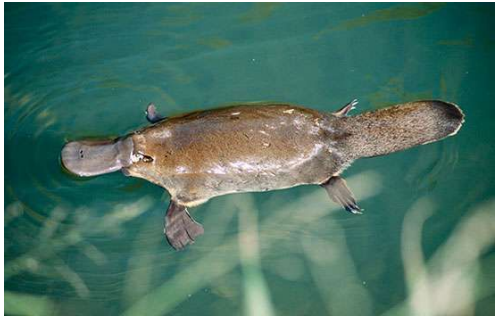
Credit Card Fraud



Model: All transactions are fraudulent.

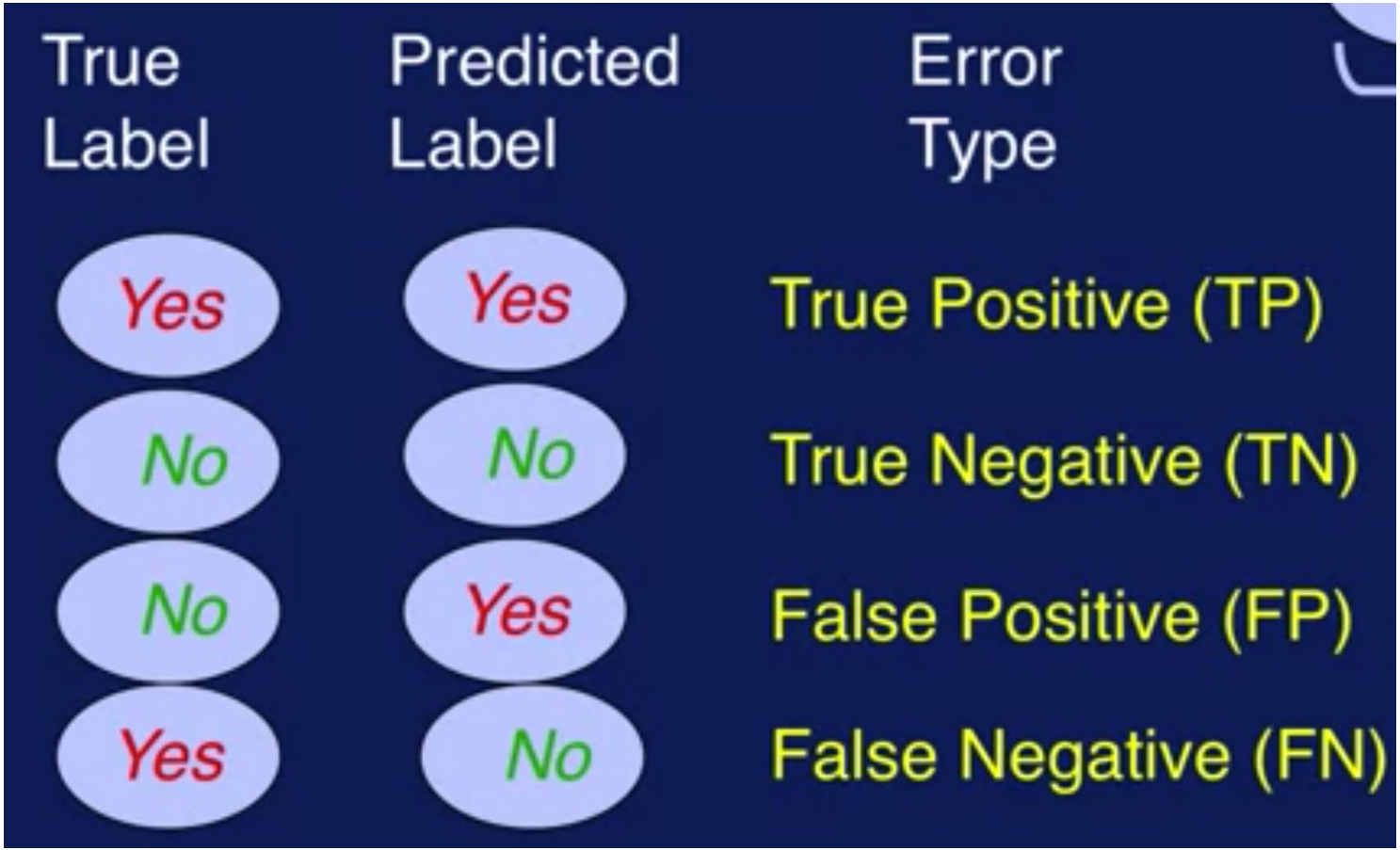
Great! Now I'm catching *all* the bad transactions!

Problem: I'm accidentally catching all the good ones!



Classic Classification Problem - “Are these animals mammals?”

Types of Classification Errors



TN

FN

FN

TP

Characteristics

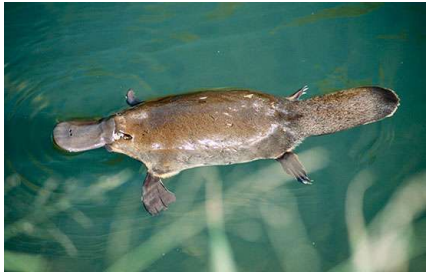
- Mammary Glands
- Fur or hair
- Three middle ear bones

TN

FP

TP

FP



Measuring Machine Learning Models

❑ Classification Metrics

- ✓ Confusion Matrix
- ✓ Accuracy
- ✓ Precision
- ✓ Recall
- ✓ F1 Score
- ✓ F-Beta Score
- ✓ Receiver Operating Characteristic Curve – ROC Curve

❑ Regression Metrics

- ✓ Mean Absolute Error
- ✓ Mean Squared Error
- ✓ R2 Score

$$\text{Accuracy Rate} = \frac{\#\text{ correct predictions}}{\#\text{ total predictions}}$$

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

$$\text{Error Rate} = \frac{\#\text{ incorrect predictions}}{\#\text{ total predictions}}$$

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

$$= 1 - \text{Accurate Rate}$$

True Label	Predicted Label
Yes	No
No	No
No	No
Yes	Yes
Yes	Yes
No	No
Yes	No
Yes	Yes
No	No
No	Yes

$$\text{Accuracy Rate} = \frac{\#\text{ correct predictions}}{\#\text{ total predictions}} = \frac{TP + TN}{TP + TN + FP + FN} = (3 + 4) / 10 = 7 / 10 = 0.7$$

$$\text{Error Rate} = \frac{\#\text{ incorrect predictions}}{\#\text{ total predictions}} = 1 - \text{Accuracy Rate} = 1 - 0.7 = 0.3$$

Limitation with Accuracy

Is this tumor cancerous?

very few positive examples

most are negative examples

Class Imbalance Problem

Limitation with Accuracy

Is this tumor cancerous?

- Say 3% of samples are cancer
- If model always predicts non-cancer
  - Accuracy = 97%

97% of samples are non-cancerous

- But no cancer cases detected!

Class Imbalance Problem

Precision and Recall : Evaluation Metrics required to find out how well our model classifies positive versus negative classes

$$\text{Precision} = \frac{TP}{TP + FP}$$

← All samples with Predicted = Yes

Denominator: Total # of samples predicted as being positive

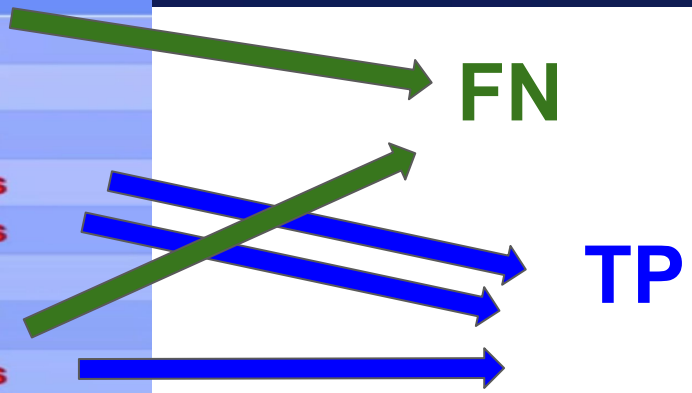
Class Imbalance Problem

Precision and Recall : Evaluation Metrics required to find out how well our model classifies positive versus negative classes

$$\text{Recall} = \frac{TP}{TP + FN}$$

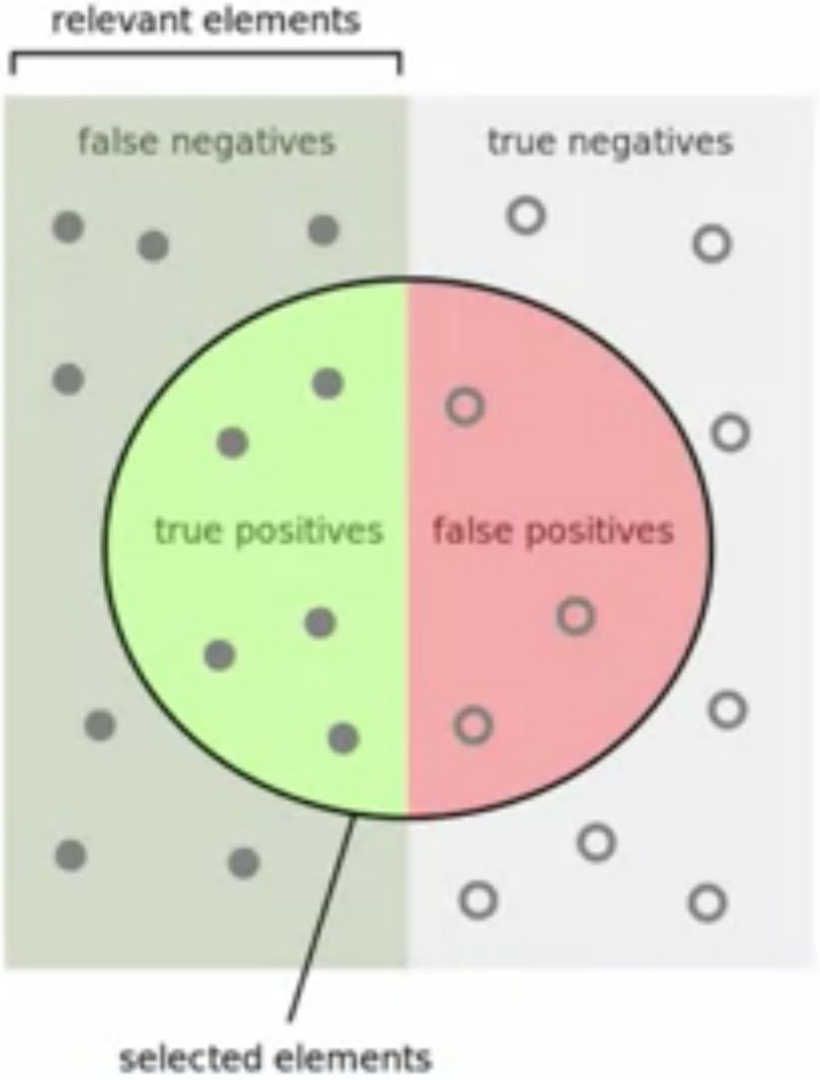
← All samples with True = Yes

True Label	Predicted Label
Yes	No
No	No
No	No
Yes	Yes
Yes	Yes
No	No
Yes	No
Yes	Yes
No	No
No	Yes



Denominator: Total # of samples belonging to the "True Class" or with "True/Actual" Label as "True"

Class Imbalance Problem

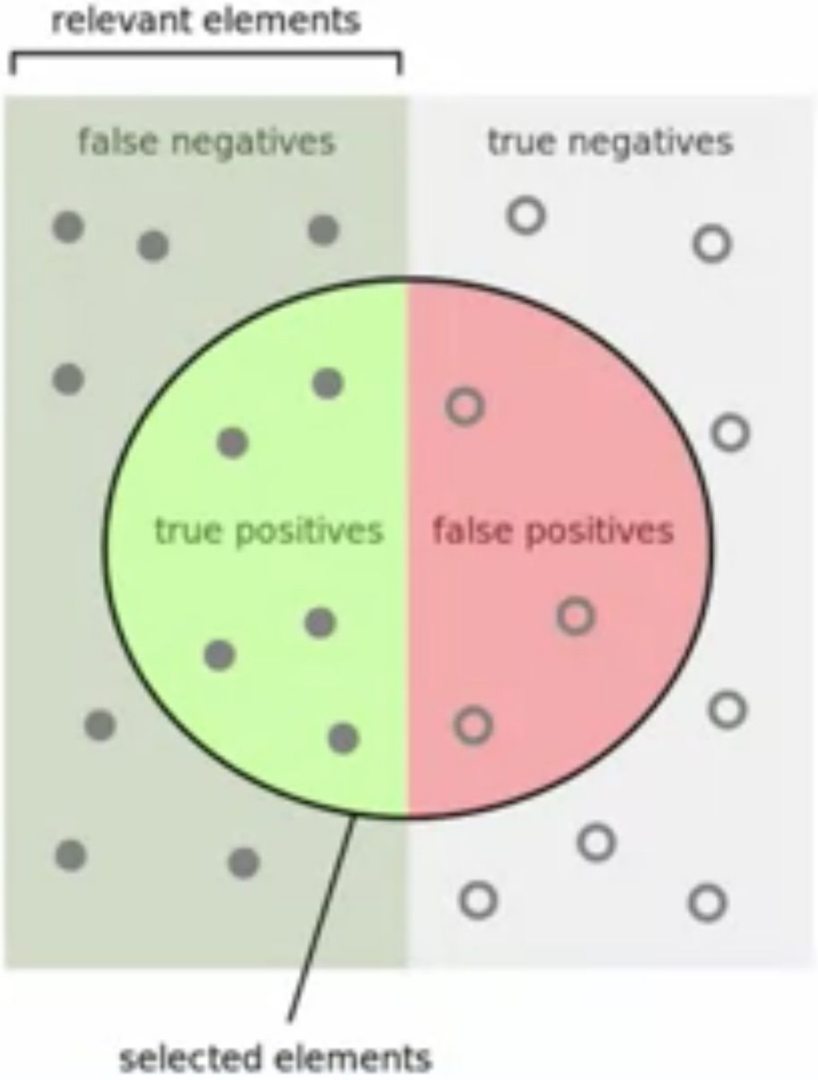


How many relevant items are selected?

Recall =

Samples correctly predicted as Positive

Samples actually Positive



How many selected items are relevant?

$$\text{Precision} = \frac{\text{Samples correctly predicted as Positive}}{\text{Samples predicted as Positive}}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

All samples with Predicted = Yes

Measure of exactness

Predicted as positive which is **actually** in the "positive" class

Calculates the number of positive classes that **the model correctly identified**

Measure of completeness

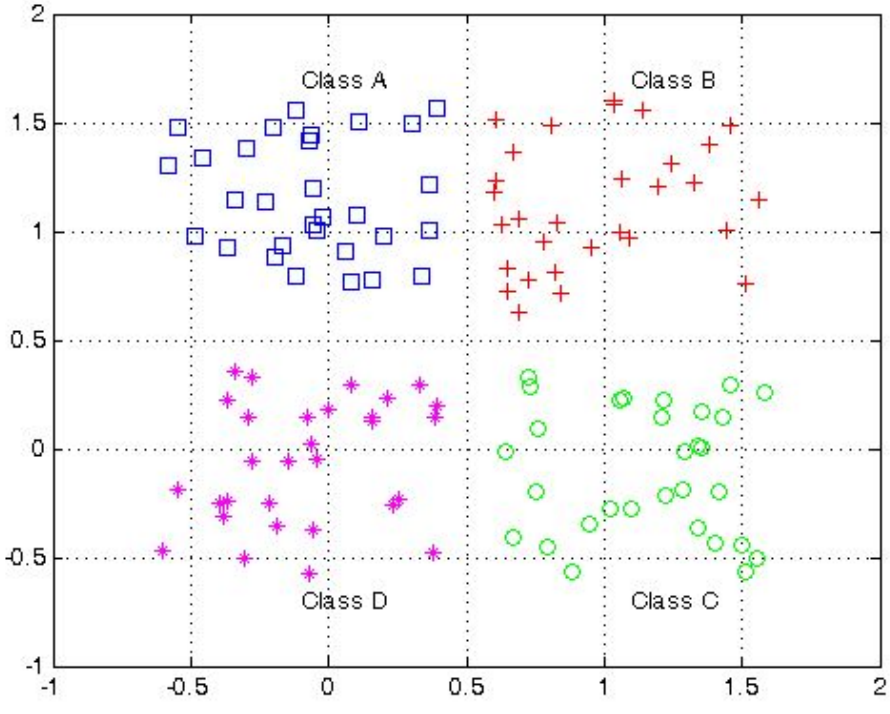
$$\text{Recall} = \frac{TP}{TP + FN}$$

All samples with True = Yes

Precision + **Recall**

- Use together
- Goal: Maximize both

For a Fixed Value of Precision, determine recall and vice-versa.



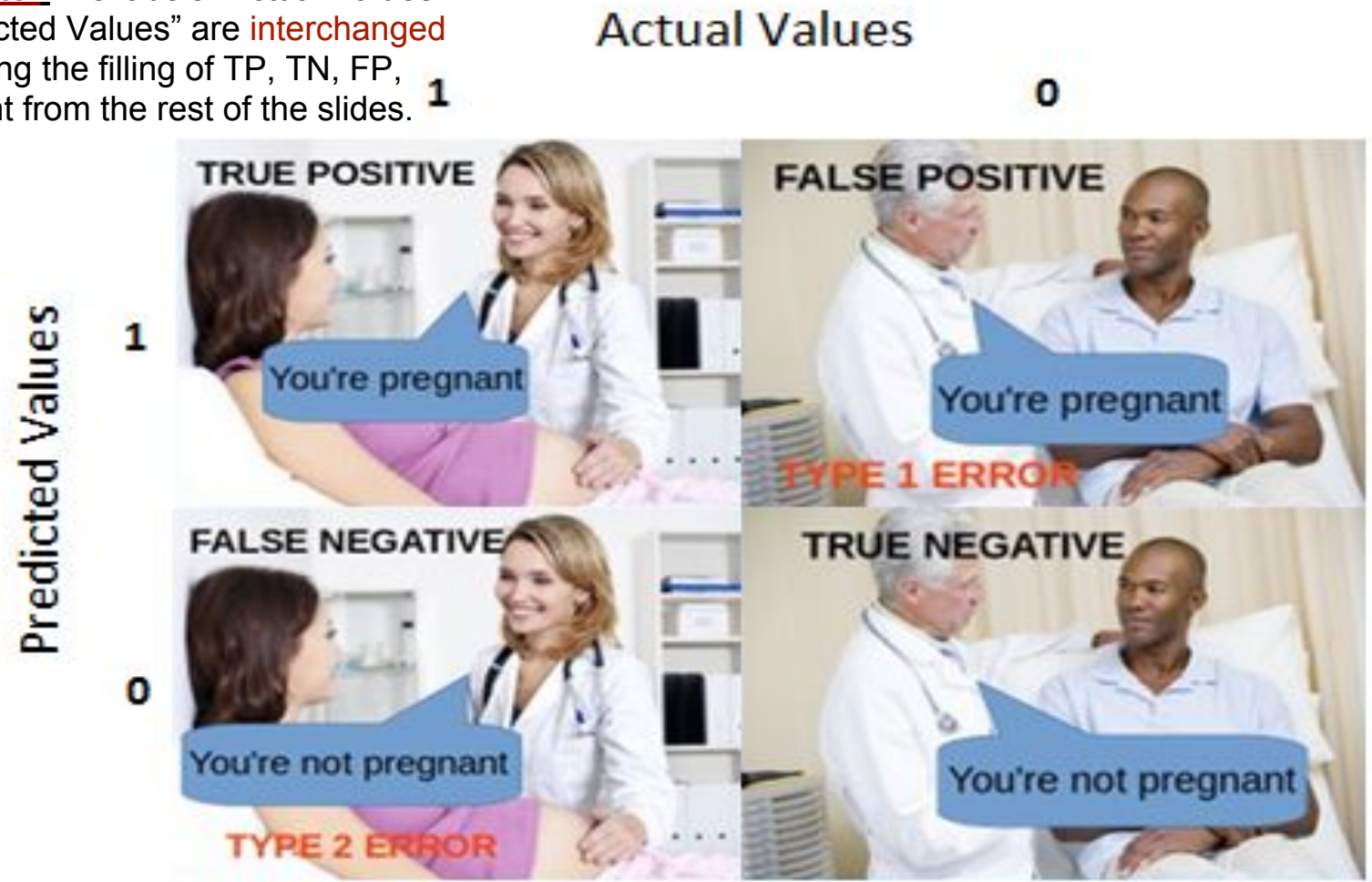
Perfect **Precision** Score of 1 for Class A samples ⇒ Every sample belonging to Class A were indeed belonging to Class A.

Missing: What about samples of Class A that were predicted / classified incorrectly?

Precision **Recall** Score of 1 ⇒ # of samples of Class A correctly labelled as Class A.

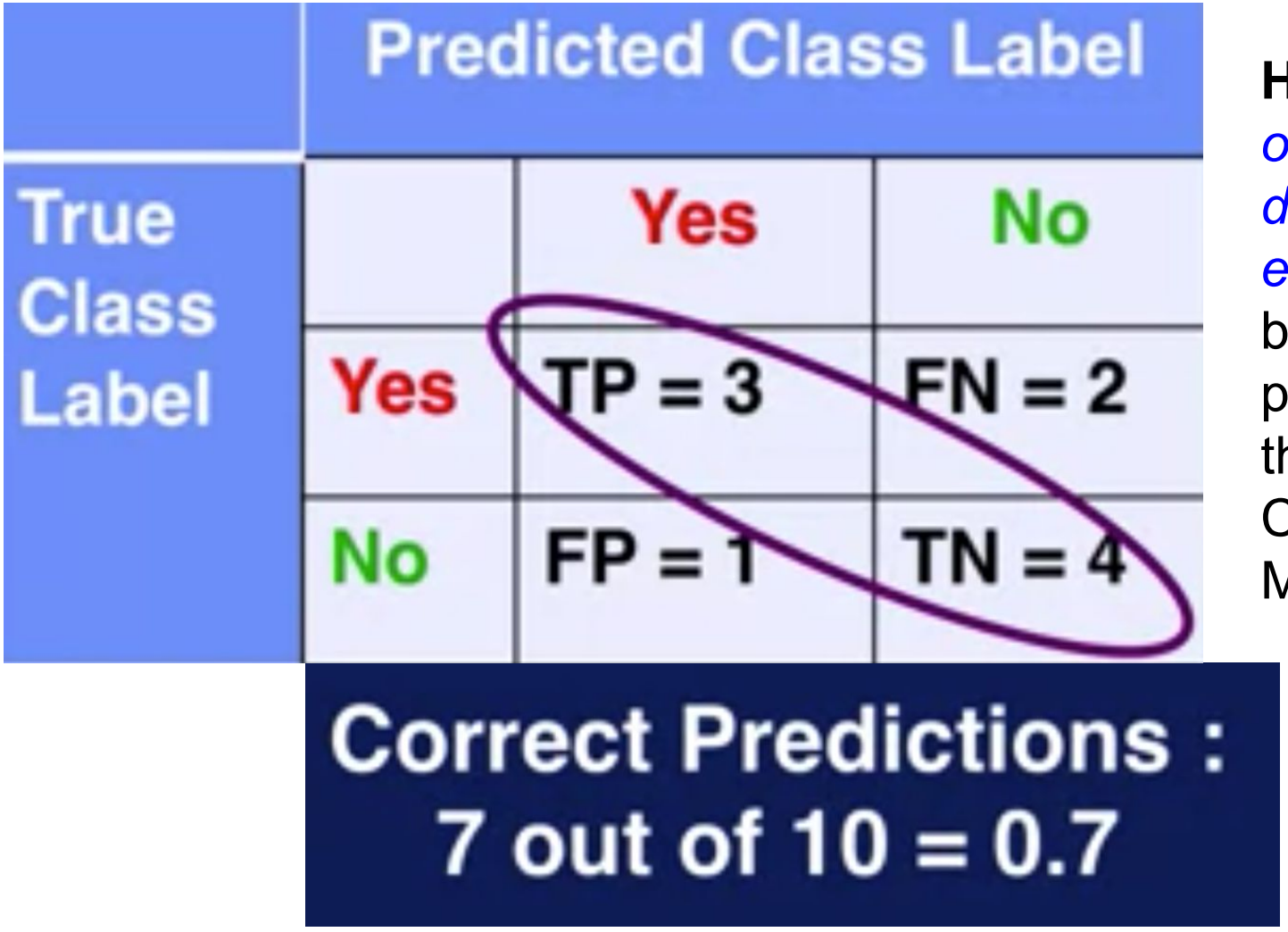
Missing: How many samples were incorrectly labelled as Class A?

Please Note: The labels “Actual Values” and “Predicted Values” are interchanged here, making the filling of TP, TN, FP, FN different from the rest of the slides.



True Label	Predicted Label
Yes	No
No	No
No	No
Yes	Yes
Yes	Yes
No	No
Yes	No
Yes	Yes
No	No
No	Yes

		Predicted Class Label	
		Yes	No
True Class Label	Yes	TP = 3	FN = 2
	No	FP = 1	TN = 4



Higher *the sum of values of diagonal elements* the better the performance of the Classification Model.

	Predicted Class Label		
True Class Label		Yes	No
	Yes	TP = 3	FN = 2
	No	FP = 1	TN = 4

Lower *the sum of values of off-diagonal elements* the better the performance of the Classification Model.

Three mis-classifications!!!

Incorrect Predictions : 3 out of 10 = 0.3

		Predicted Class Label	
		Yes	No
True Class Label	Yes	TP = 3	FN = 2
	No	FP = 1	TN = 4

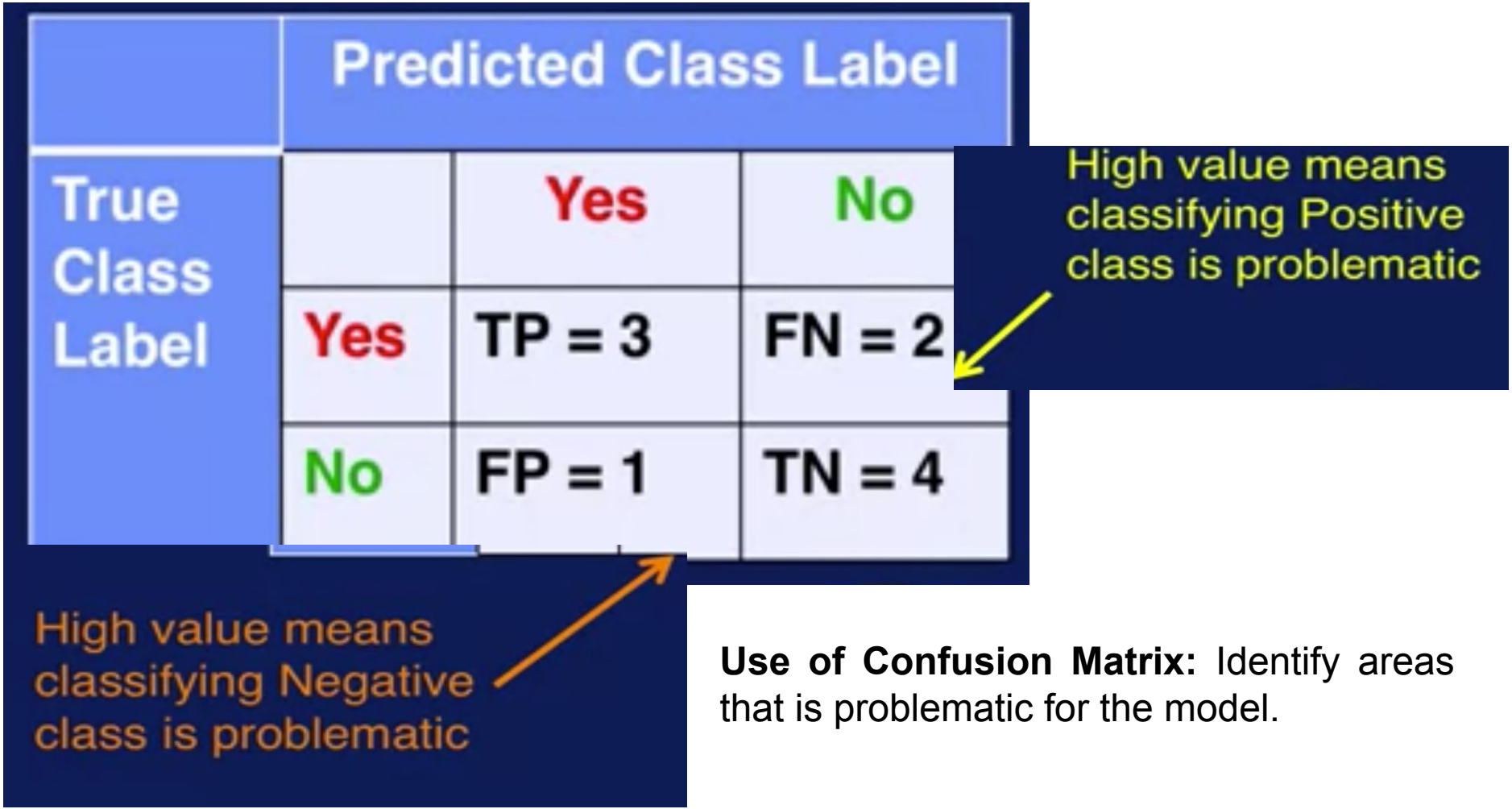
$$\text{Accuracy Rate} = \frac{\#\text{ correct predictions}}{\#\text{ total predictions}} = \frac{TP + TN}{TP + TN + FP + FN}$$

The sum of the Diagonal Elements is a measure of the **Accuracy Rate**.

		Predicted Class Label	
		Yes	No
True Class Label	Yes	TP = 3	FN = 2
	No	FP = 1	TN = 4

The sum of the Off-Diagonal Elements is a measure of the **Error Rate.**

$$\text{Error Rate} = \frac{\#\text{ incorrect predictions}}{\#\text{ total predictions}} = \frac{FN + FP}{TP + TN + FP + FN} = 1 - \text{Accurate Rate}$$



Use of Confusion Matrix: Identify areas that is problematic for the model.

CONFUSION MATRIX

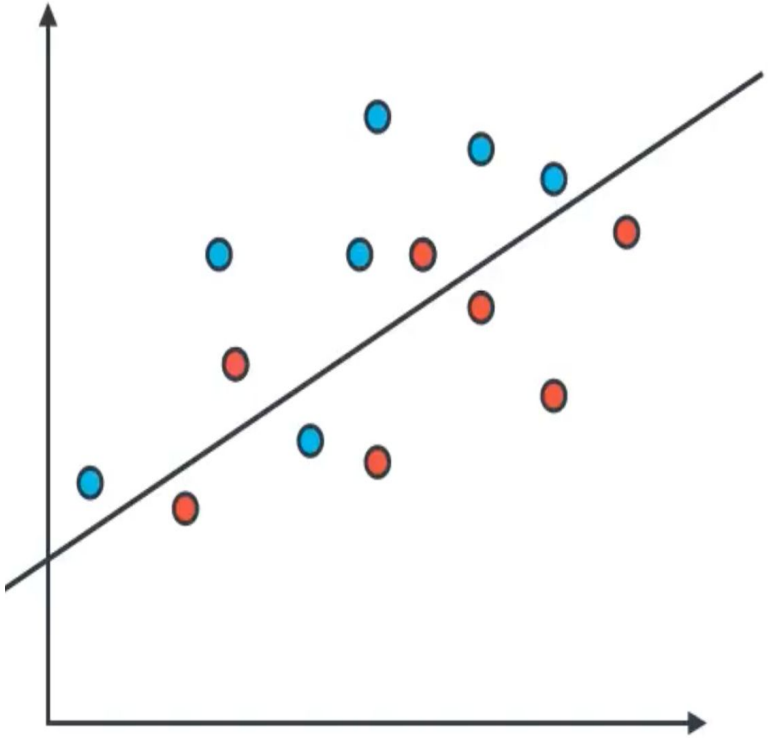
❖ **Confusion Matrix**

- ➢ a table that describes the performance of a model
- ❑ True Positive:
  - Positive point classified as positive.
- ❑ True Negative:
  - Negative point classified as negative.
- ❑ False Positive:
  - Negative point incorrectly classified as positive.
- ❑ False Negative:
  - Positive point incorrectly classified as negative.

	Guessed positive	Guessed Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

- Table used to classify the different types of errors made by a Classifier
- Values used to find the performance of a classifier (accuracy & error rate)
- Indicates what types of errors the model is making

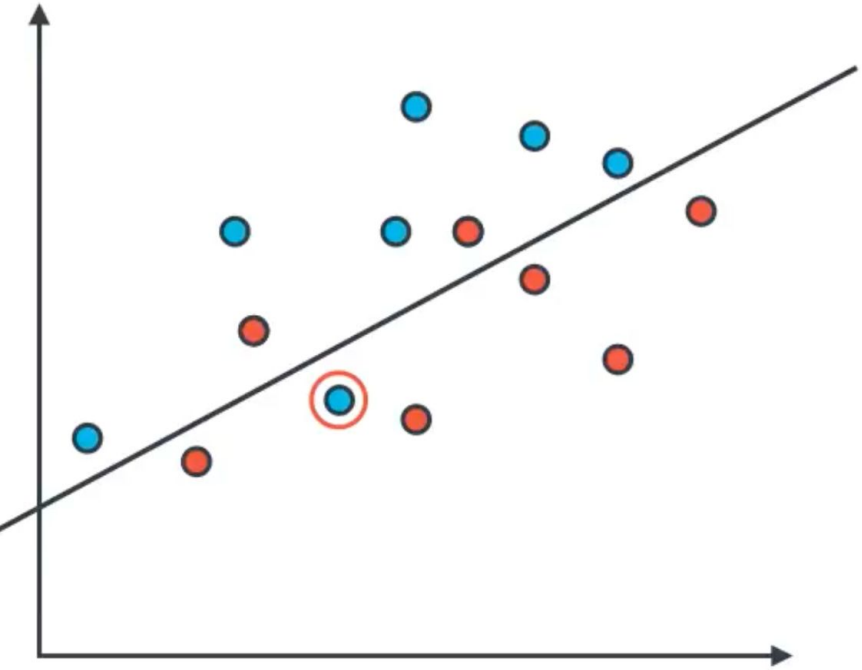
Confusion Matrix



Prediction

Data	Guessed Positive	Guessed Negative
Positive		
Negative		

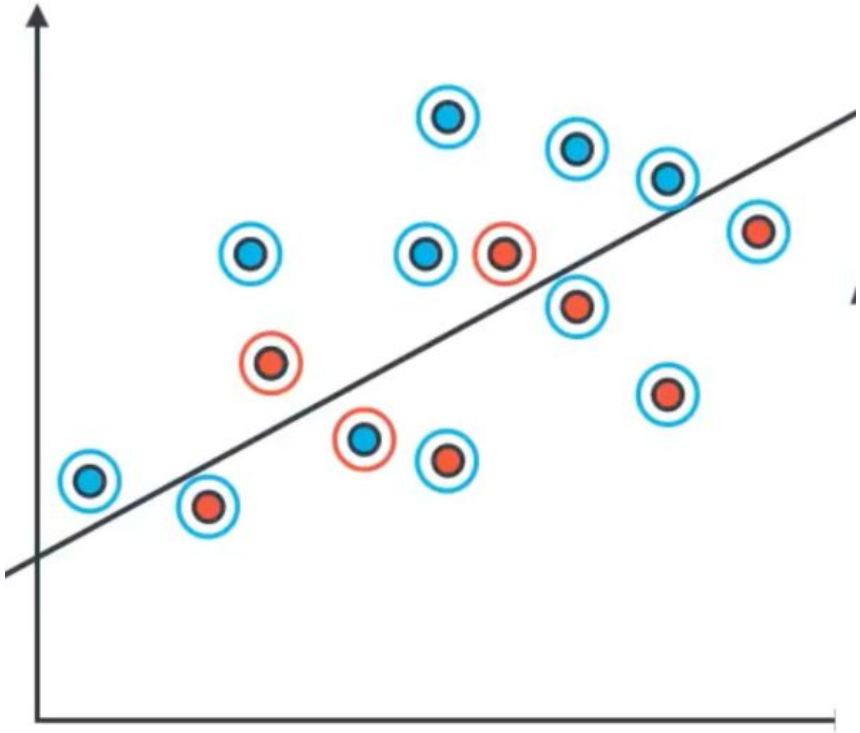
Confusion Matrix



Prediction

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

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

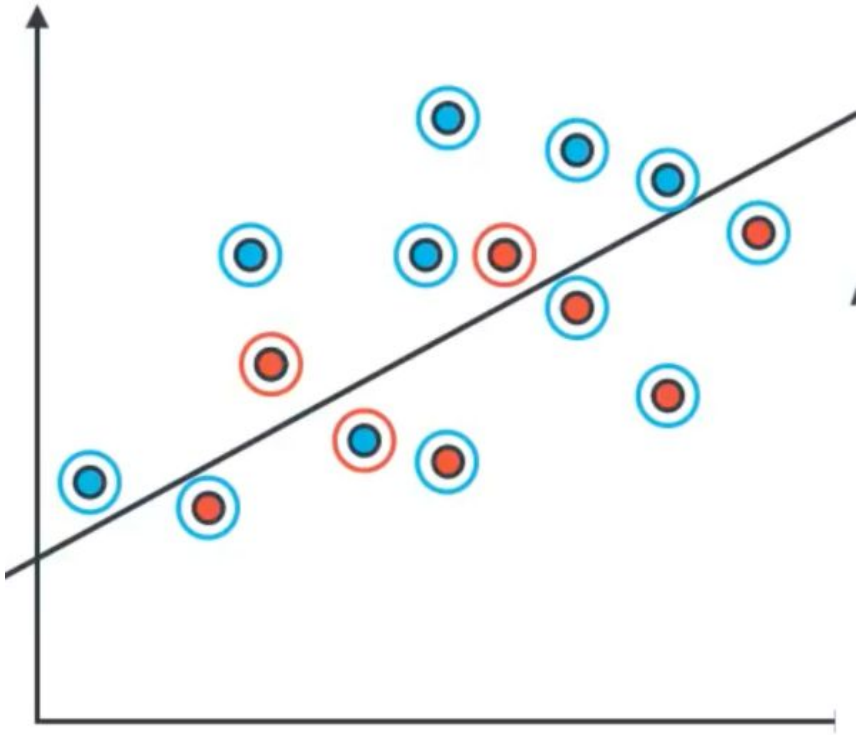


$$\text{Accuracy} = \frac{\text{Correctly Classified points}}{\text{All points}}$$

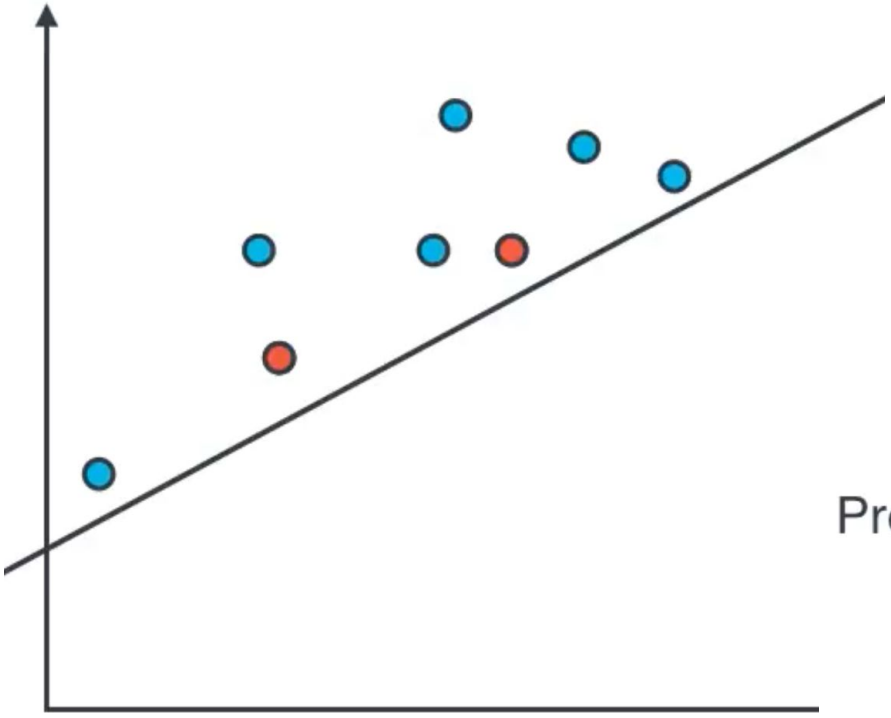
$$= \frac{11}{11 + 3}$$

$$= \frac{11}{14}$$

$$= 78.57\%$$



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



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

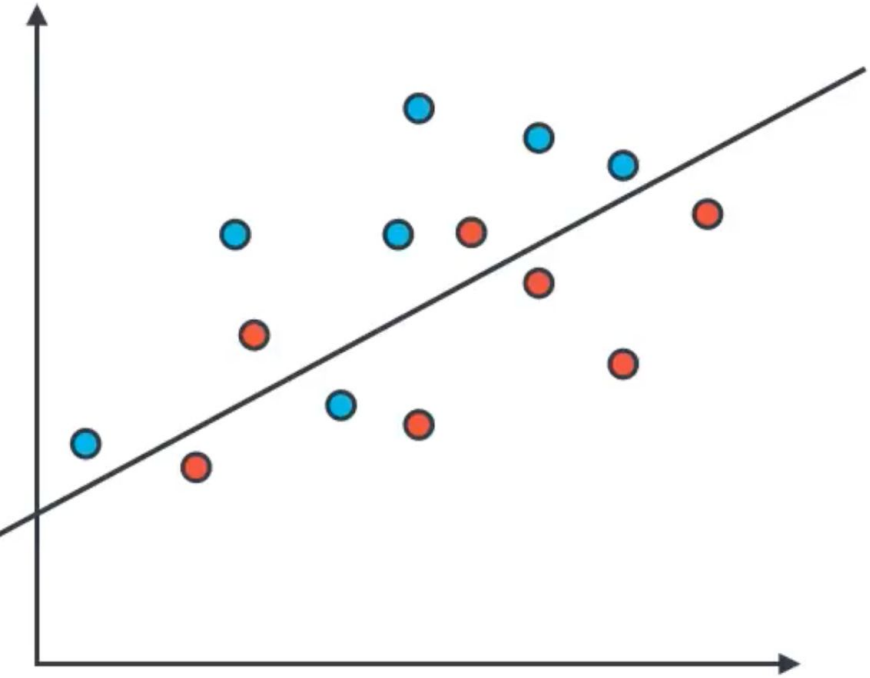
$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False Positives}}$$

$$= \frac{6}{6 + 2}$$

$$= \frac{6}{8}$$

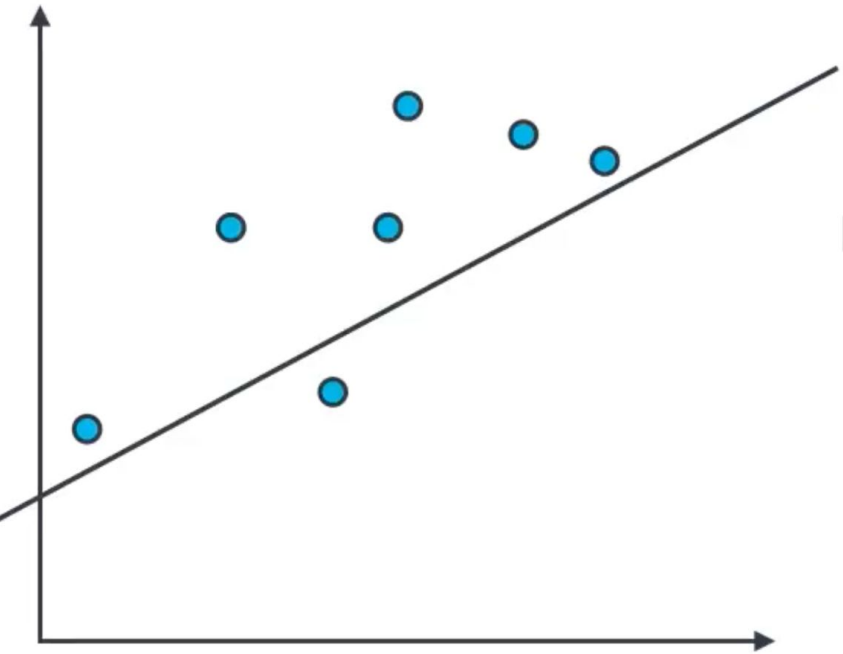
$$= 75\%$$

Recall



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

Recall



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

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

$$= \frac{6}{6 + 1}$$

$$= \frac{6}{7}$$

$$= 85.7\%$$

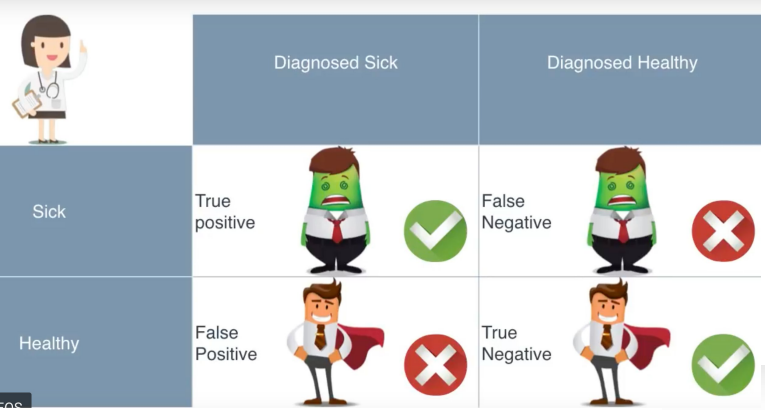
Medical Model

Healthy

Sick

Spam Classifier Model

Not spam



$$\text{Accuracy} = \frac{1{,}000 + 8{,}000}{10{,}000} = 90\%$$

10,000 Patients

Confusion Matrix

Diagnosis

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

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

$$\text{Accuracy} = \frac{100 + 700}{1000} = 80\%$$

1,000
e-mails

Confusion Matrix

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

	Diagnosed Sick	Diagnosed Healthy
Sick		False Negative  
Healthy	False Positive  	

	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

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 Recall Model

High Precision Model

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?

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

Diagnosis

Patients	Diagnosed sick	Diagnosed Healthy
Sick	1000	200
Is Healthy	600	9000

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

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?

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?

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

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?

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

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?

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?

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

Precision and Recall

Medical Model

Precision: 55.7%

Recall: 83.3%

Spam Detector

Precision: 76.9%

Recall: 37%