Data Science and Artificial Intelligence

Machine Learning

Classification

Lecture No. 6

By- SIDDHARTH SABHARWAL SIR

Recap of Previous Lecture

- **Topic** — Confusion matrix
- **Topic** — ROC, AUC
- **Topic** — TPR, FPR
- **Topic** — Sensitivity | specificity.
- **Topic**

Topics to be Covered

- **Topic** Precision / Recall
- **Topic** Time & space Complexity
- **Topic** Questions
- **Topic**
- **Topic**

Correct daily plan of the daily routine

Confusion matrix...

TPR FPR

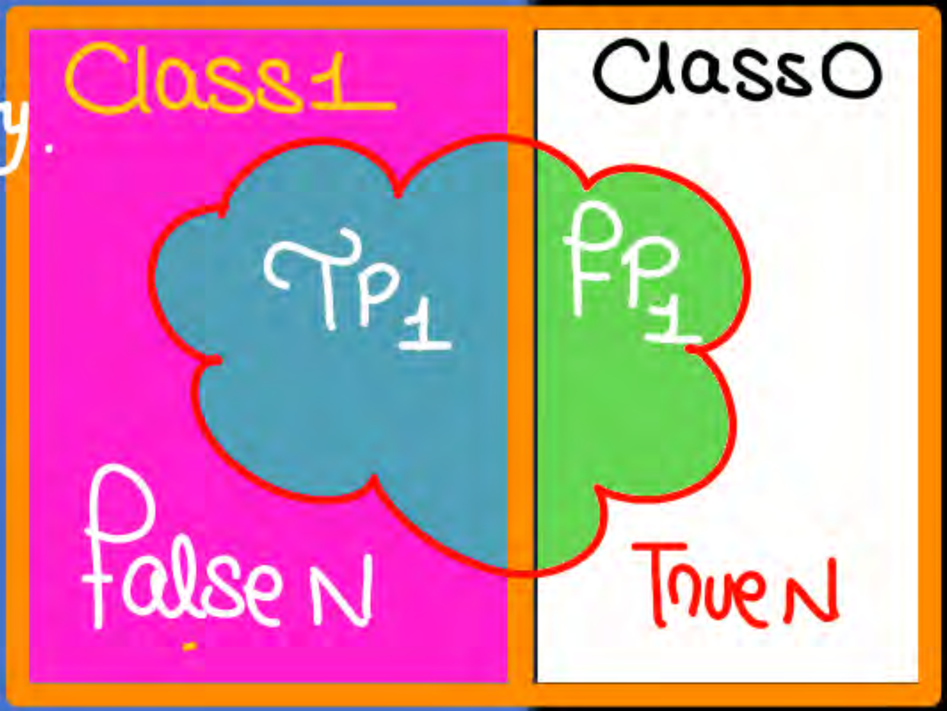
Sensitivity

TPR = TP/actual P

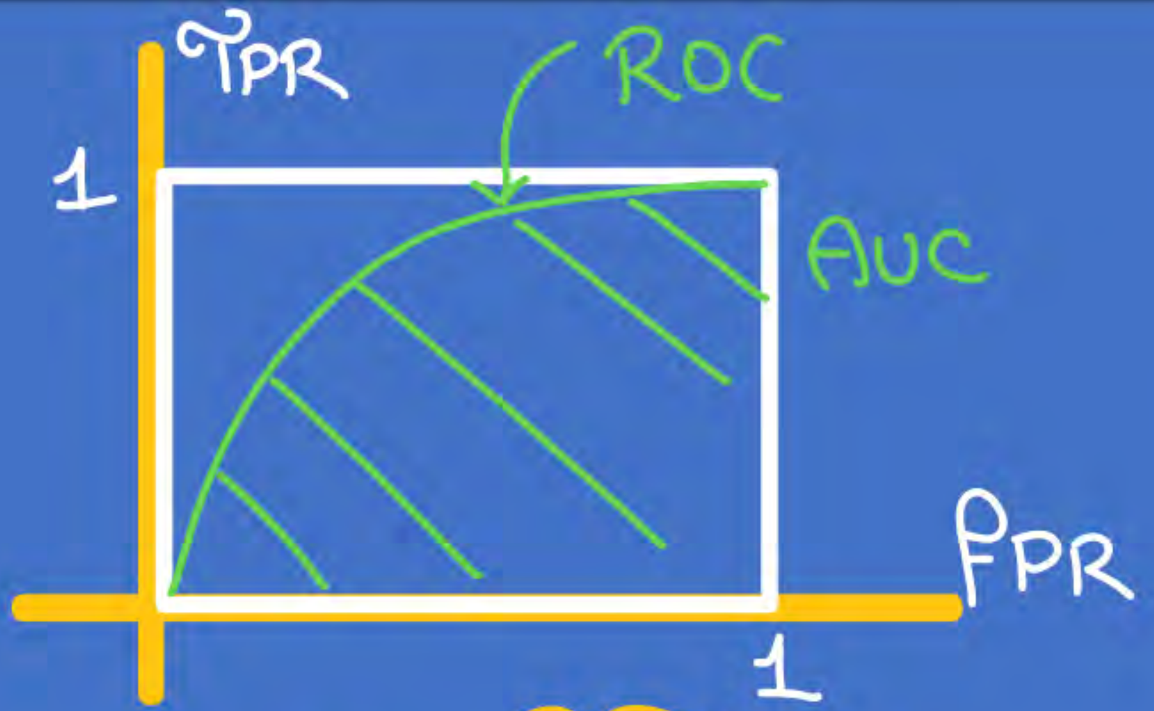
FPR = FP/actual N

1-Specificity

	Actual 1	Actual 0
Predicted 1	TP	FP
0	FN	TN



TPR and FPR done



Why only TPR and FPR ? → we want TPR → 1
FPR → 0

REVISION

- AUC = always b/w 0 → 1
  - ↙ Bekar Classifier
  - ↘ ideal class.

If TPR = 1 ⇒ TP = Actual P → FN = 0

FPR = 0 ⇒ FP = 0 → TN = Actual N

So if we Improve TPR and dec FPR, then automatically the Performance on Neg class Improve

What is Specificity and ~~selectivity~~... Sensitivity

done

accuracy on N class

$\rightarrow \frac{TN}{\text{Actual } N}$

$\rightarrow$ accuracy on P class

$\frac{TP}{\text{Actual } P}$

What is ROC and AUC?

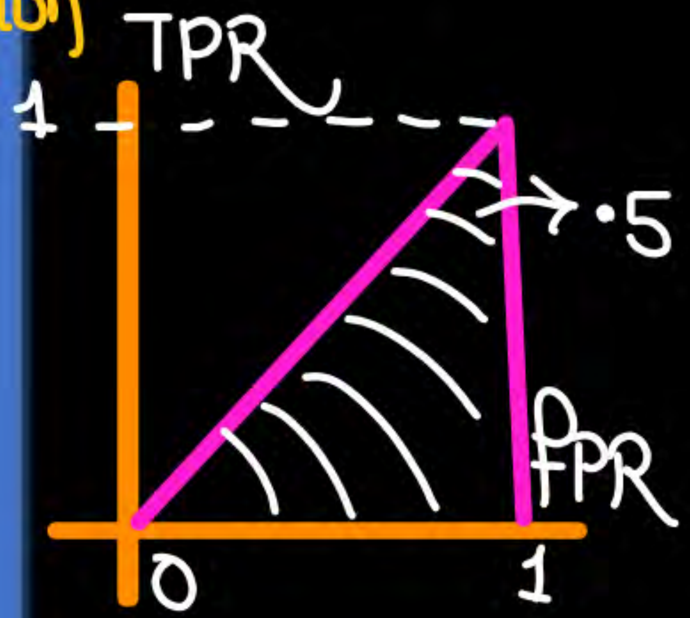
Q) If AUC for any classifier is 0.5

→ a) good

b) equivalant to Random classification

c) No Comment.

So classifier TPR = FPR



generally in any data 2 class →

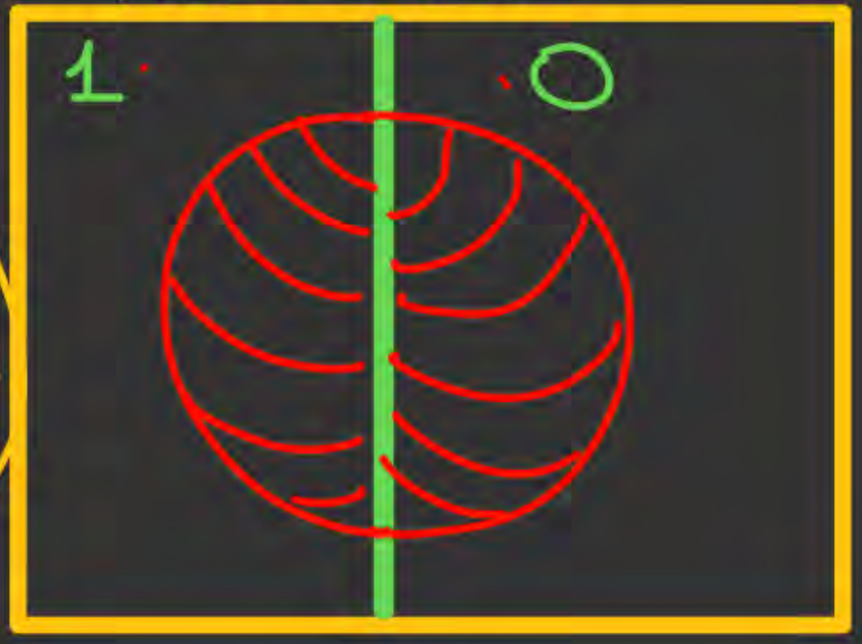
class 0 → 50%

class 1 → 50%

So TPR = FPR

→ No of Correct 1 Predicted = No of wrong 1 Predicted

Actual class



Recall and Precision

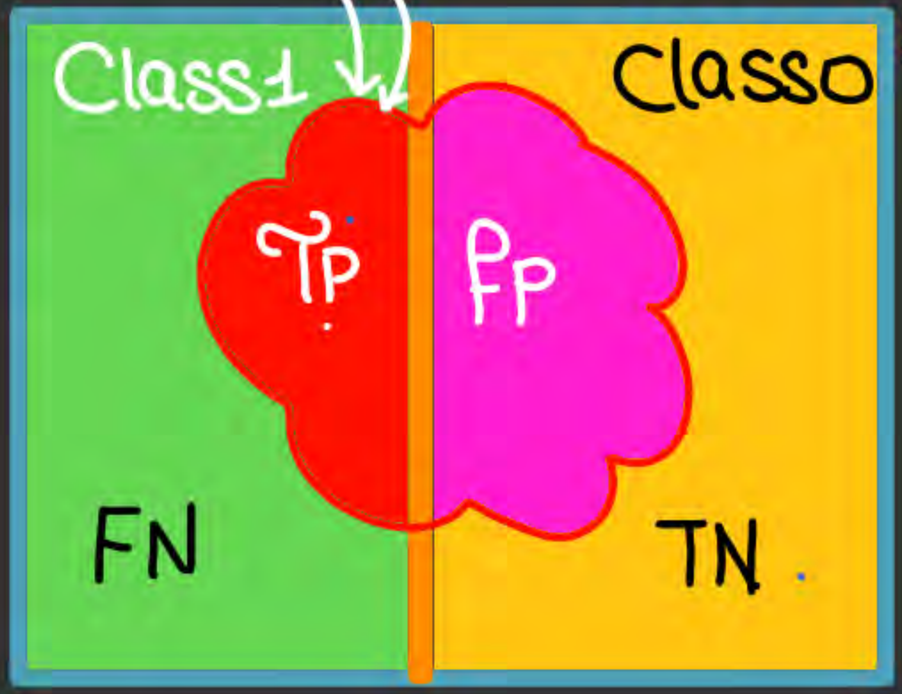
Recall ↔ TPR

TPR

Sensitivity

Recall → $\frac{TP}{\text{Actual } P}$

Precision → $\frac{TP}{TP+FP} = \left(\frac{TP}{\text{Total No of Predicted } P}\right)$



→ Accuracy → $\frac{TP+TN}{\text{Total No of Points}}$

- So we want Recall and precision shd be equal to '1' ideally

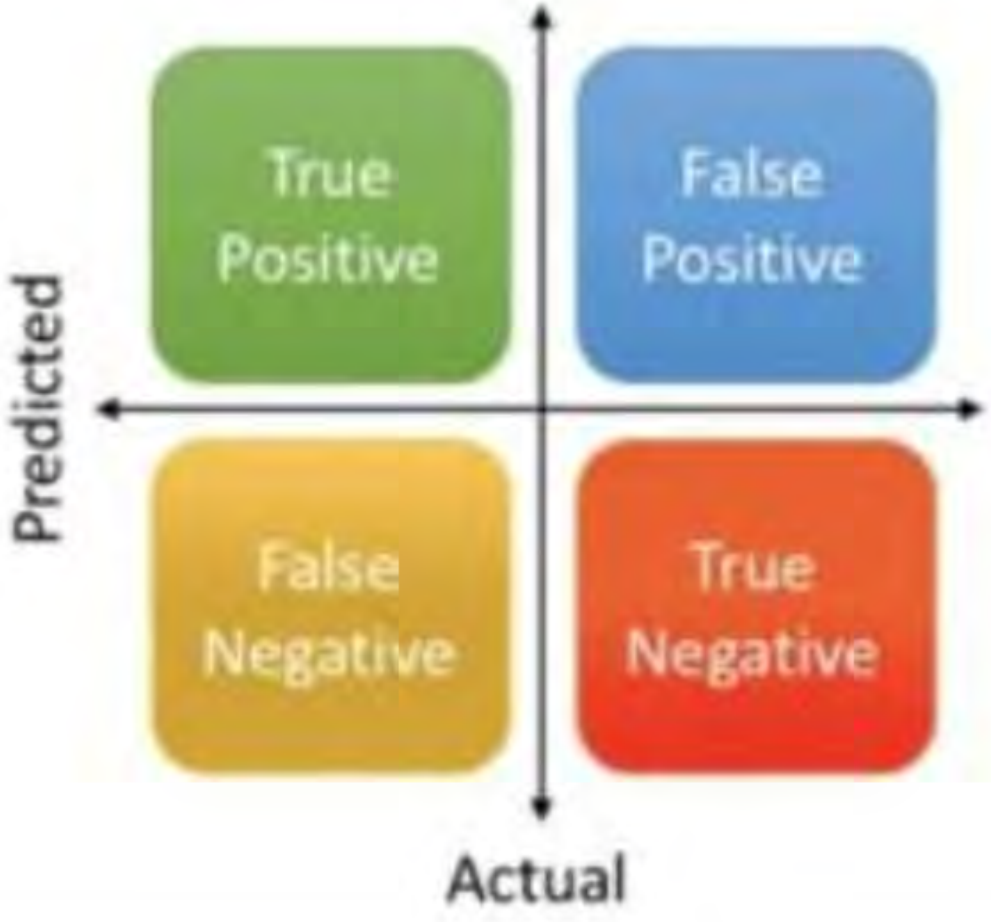
Linear Classification

What is Recall and Precision

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

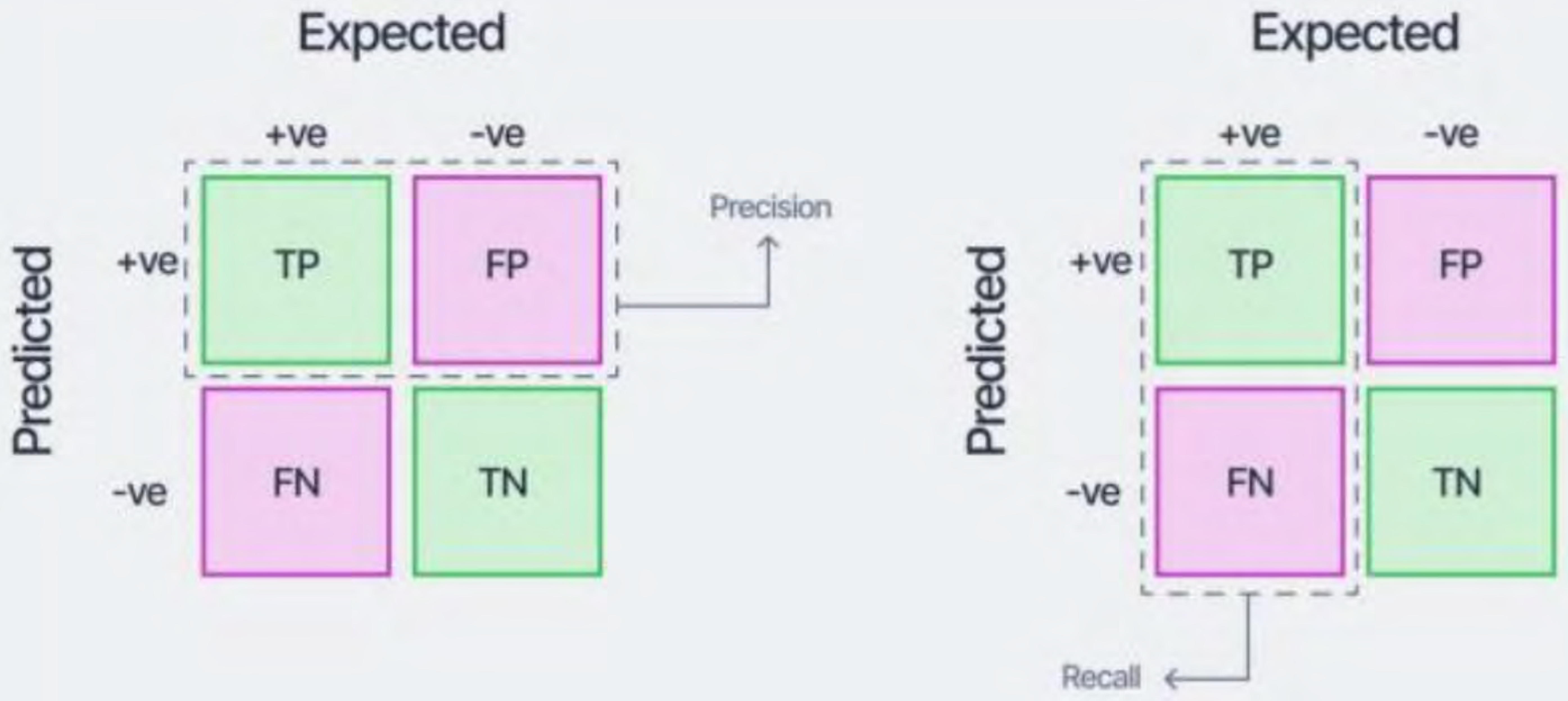
$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

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



Linear Classification

What is Recall and Precision



$\Uparrow$ more Imp.

Recall vs Precision

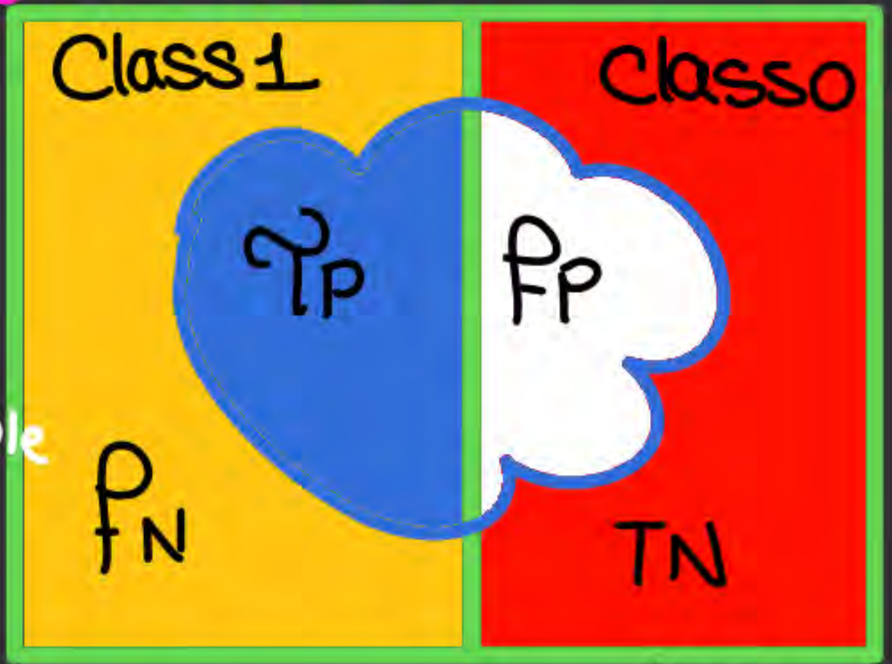
$$\frac{TP}{TP+FN}$$

- Video was good but algo decline

$$\frac{TP}{TP+FP}$$

Video has objectionationable but algo allowed it

QYT alogorithm to check Videos and allow (Positive) or not allow (Neg) the video to <18 year children



- In this case FP is very dangerous, it has to be zero. FN can be non zero

Precision shd be high.

Q. If the patient is Covid +ve (P)
and if patient is Covid -ve (N)

More Imp

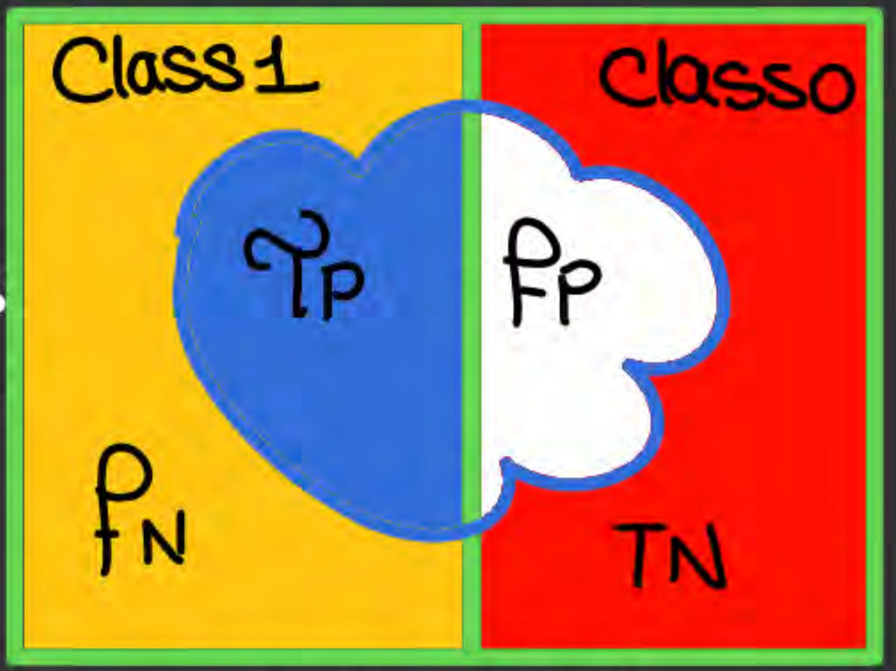
Recall Vs Precision

Recall: $\dfrac{TP}{TP + FN}$

- Patient was Covid +ve but algo say false No

Precision: $\dfrac{TP}{TP + FP}$

Patient was not Covid +ve but doctor say false P



• + • ⇒ Predicted P
• + • ⇒ Predicted N
• + • ⇒ Actual P
• + • ⇒ Actual N

→ FN and FP both shd be zero
→ FN has to be zero

True/Actual → True/Predicted

What is macro recall and precision

⇛ Since we have 2 classes

$$Recall_P = \frac{TP}{Actual\,P} \qquad Recall_N = \frac{TN}{Actual\,N}$$

$$Precision_P = \frac{TP}{Predicted\,P} \qquad Precision_N = \frac{TN}{Predicted\,N}.$$

$$macro\,Recall = \frac{1}{2}(Recall_P + Recall_N)$$

$$macro\,Precision = \frac{1}{2}(Precision_P + Precision_N)$$

alone

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.

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.

F1 Score → Harmonic mean of Recall & Precision.

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

$$F_1\ \text{score} \Rightarrow 2 \times \frac{P \times R}{P + R}$$

→ Since max value of

$P = 1$

$R = 1$

$$\Rightarrow F_1\ \text{score}_{max} = 2 \times \frac{1 \times 1}{1 + 1}$$

$$\Rightarrow 1$$

- The harmonic mean is more effected by the smaller value
- If P is low, R is large then also F1 score will be low
- So a good classifier shd have a large F1 score which is possible only when both R and P are large

Practise

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

- Accuracy
- Stability
- Connectivity
- Comparativity

Why gradient descent is not used in logistic Reg??

→ min of a fxn

→ Convex nature cost fxn

Linear regr.

$\sum(y-\hat{y})^2$

$\beta^{new} = \beta^{old} - \eta \frac{\partial L}{\partial \beta}$

Max $\prod (P_i)^{y_i}(1-P_i)^{1-y_i}$

Linear Classification

Practise

From the above Table

N= 200

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

$TPR \Rightarrow \frac{TP}{Actual\,P} \Rightarrow 125/130$

$Recall \Rightarrow TPR = \frac{125}{130}$

$Accuracy \Rightarrow \frac{TP+TN}{Total} \Rightarrow \frac{125+60}{200}$

$FPR \Rightarrow \frac{FP}{Actual\,N} \Rightarrow \frac{10}{70}$

$Precision \Rightarrow \frac{TP}{Total\,PRedicted} \Rightarrow \frac{125}{135}$

Practise

For the below confusion matrix, what is the recall?

Predicted

Actual	Not 5	5
Not 5	53272	1307
5	1077	4344

- 0.7
- 0.8
- 0.9
- 0.95

Recall ⇒ TPR

$\Rightarrow \frac{TP}{Actual\ P}$

$\Rightarrow \frac{4344}{1077+4344}$

$\Rightarrow 0.8$

Practise

For the below confusion matrix, what is the precision?

Predicted P

Actual P

	Not 5	5
Not 5	53272	1307
5	1077	4344

- 0.73
- 0.76 ✓
- 0.78
- 0.82

$$P = \frac{TP}{\text{Total Predicted P}}$$

$$= \frac{4344}{4344+1307}$$

$$= .76$$

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

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

$R = .8$

$P = .76$

$F_1 \text{score} = 2 \times \frac{.8 \times .76}{.8 + .76} = (.77948)$

$\approx .78$

What is F-1 Score

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

done

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

Question 1: Accuracy is simply a ratio of correctly predicted observations to the total observations. From the above confusion matrix, how would you define Accuracy?

(A) Accuracy = (FP+FN)/(TP+FN+FP+TN)

(B) Accuracy = (TP+TN)/(TP+FN+FP+TN)

(C) Accuracy = (TP+FN)/(TP+FN+FP+TN)

(D) Accuracy = (FP+TN)/(TP+FN+FP+TN)

Question 2: From the above confusion matrix, how would you define Error?

(A) Error = (FP+FN)/(TP+FN+FP+TN)

(B) Error = (TP+TN)/(TP+FN+FP+TN)

(C) Error = (TP+FN)/(TP+FN+FP+TN)

(D) Error = (FP+TN)/(TP+FN+FP+TN)

Question 7: Which statements about the Type-I and Type-II errors are correct? (Select two)

(A) Type-I error = FPR

(B) Type-II error = FNR

(C) Type-I error = FNR

(D) Type-II error = FPR

$\rightarrow$ Type 1 error $\Rightarrow$ FPR $\Rightarrow \frac{FP}{\text{Total Actual N}}$

$\rightarrow$ Type 2 error $\Rightarrow$ FNR $\Rightarrow \frac{FN}{\text{Total Actual P}}$

Question 8: The F1 score is the harmonic mean of Precision and Recall. What's the correct formula for the F1 score?

(A) F1 score = (2*Precision*Recall)/(Precision + Recall)

(B) F1 score = (2*Precision*TPR)/(Precision + TPR)

(C) F1 score = (2*Precision*Sensitivity)/(Precision + Sensitivity)

(D) All of the above

Question 9: Which statement about the Receiver Operating Characteristic (ROC)-(Area Under the Curve) AUC Curve is correct?

(A) ROC is a probability curve that plots the true positive rate (sensitivity or recall) against the false positive rate (1 — specificity) at various thresholds.

(B) AUC is the area under the ROC curve. If the AUC is high (close to 1), the model is better at distinguishing between positive and negative classes.

(C) If AUC = 0.5, it represents a model that is no better than random.

(D) All of the above.

What is the purpose of the ROC curve in logistic regression?

A) To assess the goodness of fit of the model.
B) To evaluate the impact of outliers on the model.
C) To visualize the relationships between independent variables.
D) To assess the trade-off between sensitivity and specificity.

What statistic is commonly used to evaluate the overall performance of a logistic regression model?

A) R-squared (R^2)
B) Mean squared error (MSE)
C) AUC-ROC (Area Under the Receiver Operating Characteristic Curve)
D) Pearson correlation coefficient (r)

Time and Space Complexity of Linear Classification

- Linear regression
  Ridge Regression

$D+1 \Downarrow K$

- Time Complexity
  $O(K^3 + NK^2)$
- Testing $O(K)$
- Space $\Rightarrow O(K)$

Linear classification $\Rightarrow$

Cost fxn $\Rightarrow \max \sum_{i=1}^{N} y_i x_i \beta$

$y_i\left[\beta_0 + \beta_1 x_i^1 + \dots + \beta_D x_i^D\right]$

$(D+1)$ Multiplication

$+1$

$\Rightarrow (D+2)$

N data points

$K = (D+1)$, β values

$\Rightarrow$ Overall Mult $\Rightarrow N(D+2)$

$\Rightarrow N(K+1)$

$\approx NK$

$\Rightarrow$ The time Complexity $\Rightarrow O(NK)$

Time and Space Complexity of Linear Classification

Testing ⇒ Time Complexity

we have β

Only Find $(x_i \beta) = (\beta_0 + \beta_1 x_i^1 + \ldots \beta_D x_i^D)$

⇒ $(D+1)$ multiplication

⇒ $O(K)$ Complexity

Space Complexity ⇒ we have to store $D+1$ β's

⇒ $O(K)$.

Nearest Neighbour Method

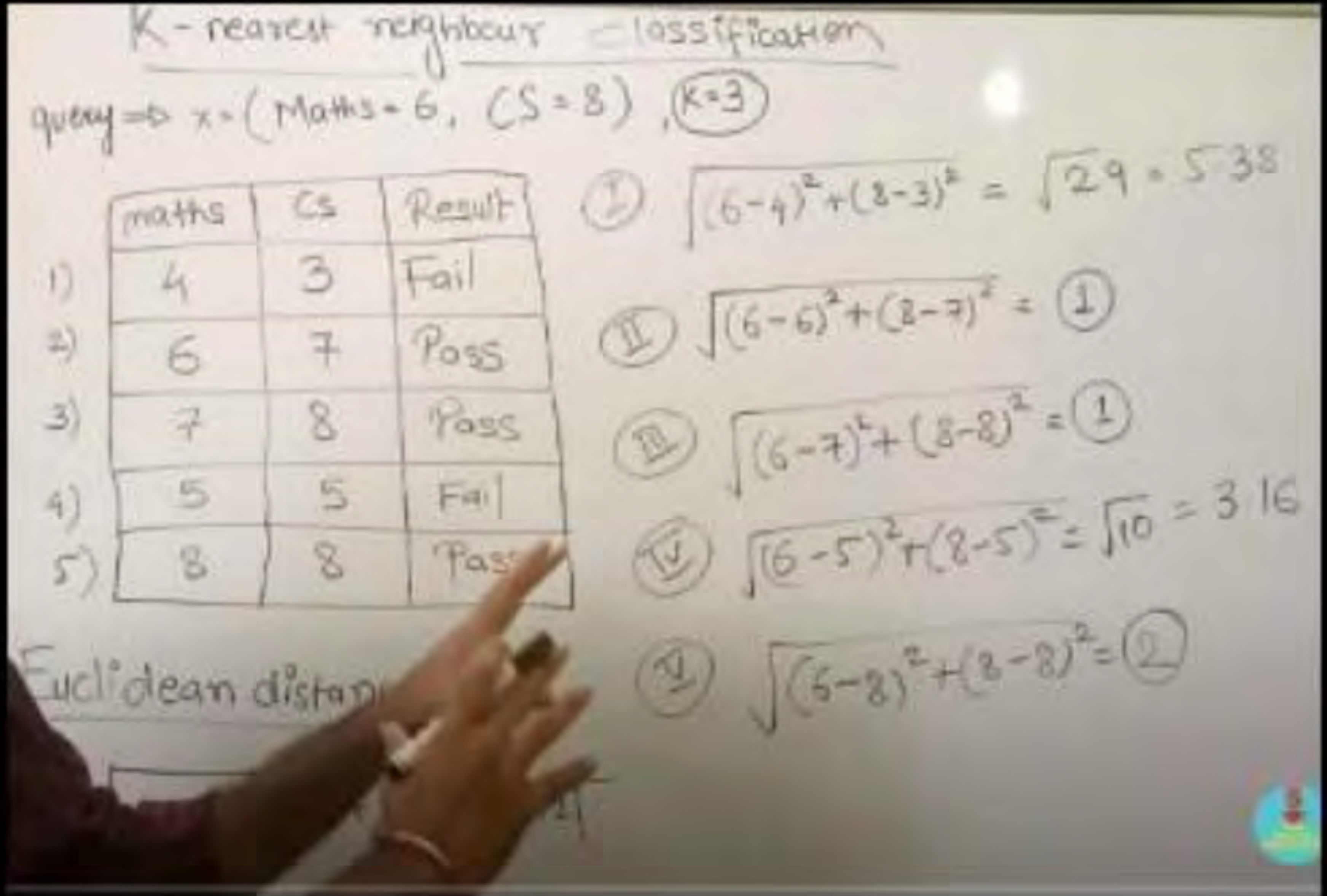
K-NN method of Supervised Learning

The general flow of the Supervised Learning Algorithms

Nearest Neighbour Method

K-NN method of Supervised Learning

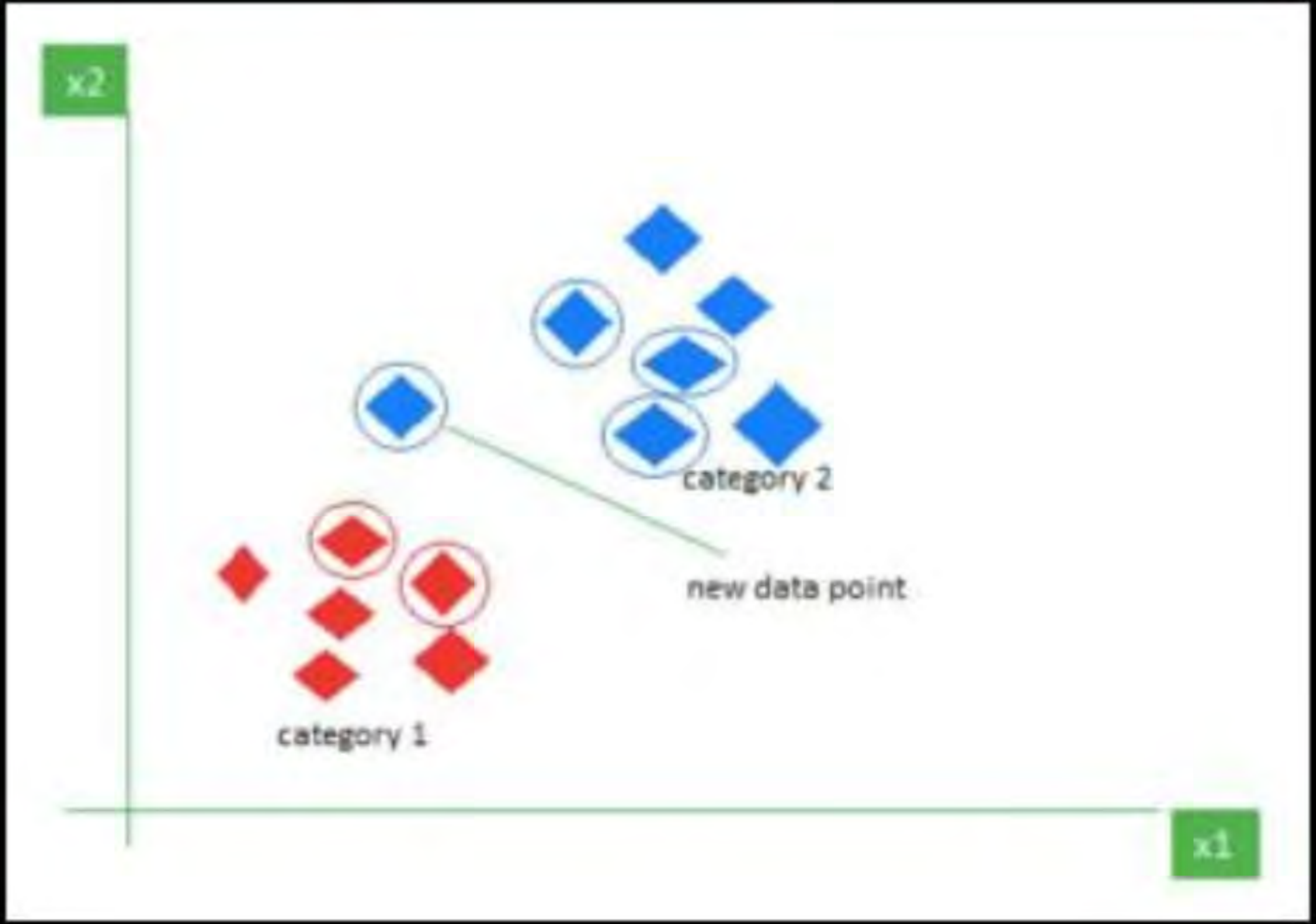
Lets understand NN with a simplest example...



Nearest Neighbour Method

K-NN method of Supervised Learning

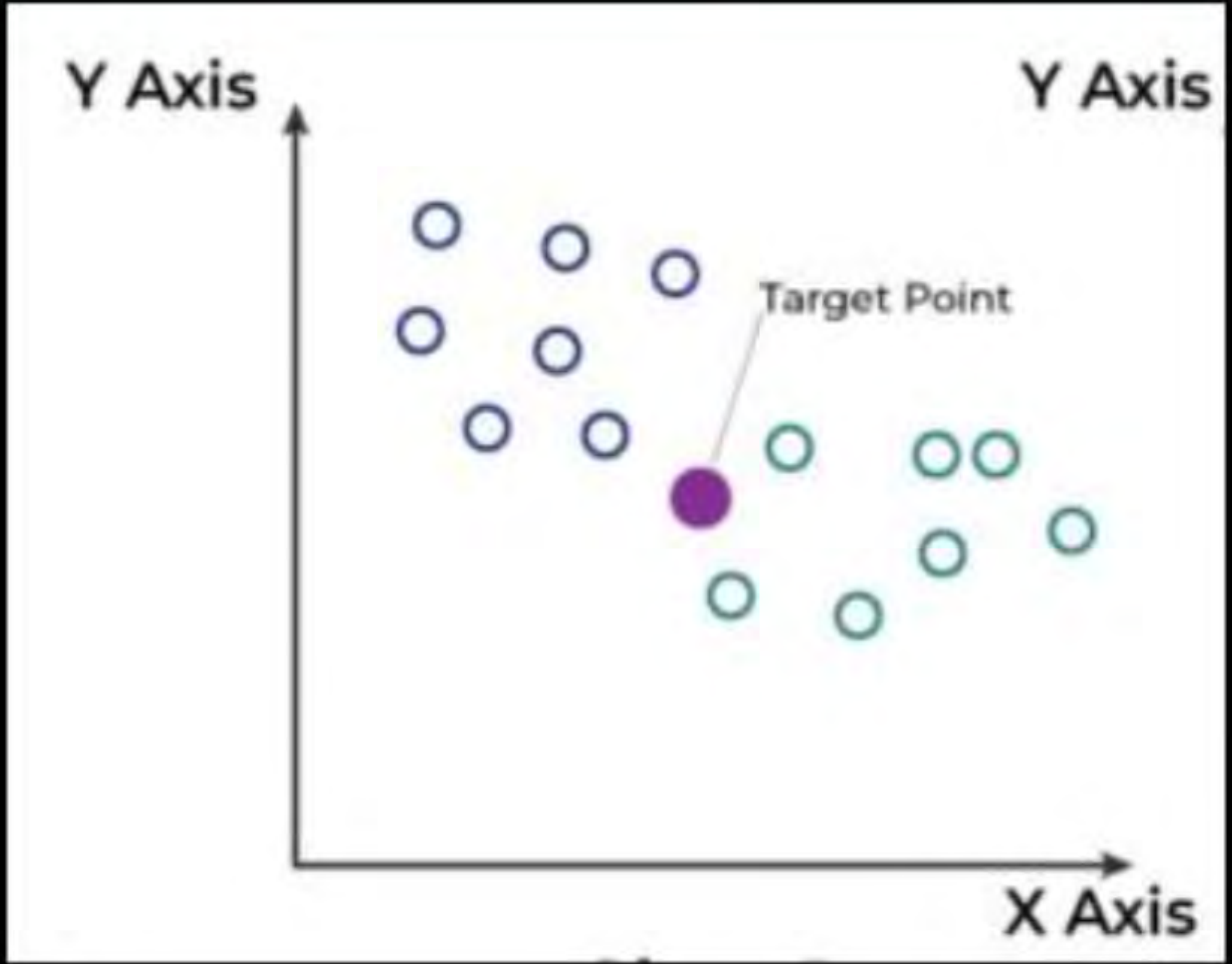
Lets understand NN with a simplest example...



Nearest Neighbour Method

K-NN method of Supervised Learning

Lets understand NN with a simplest example...



Nearest Neighbour Method

K-NN method of Supervised Learning

The general flow of the Supervised Learning Algorithms

Step 1

Process the data, train the model, find the parameters

Step 2

Use the parameters for regression and classification

This is one of the simplest supervised Learning model

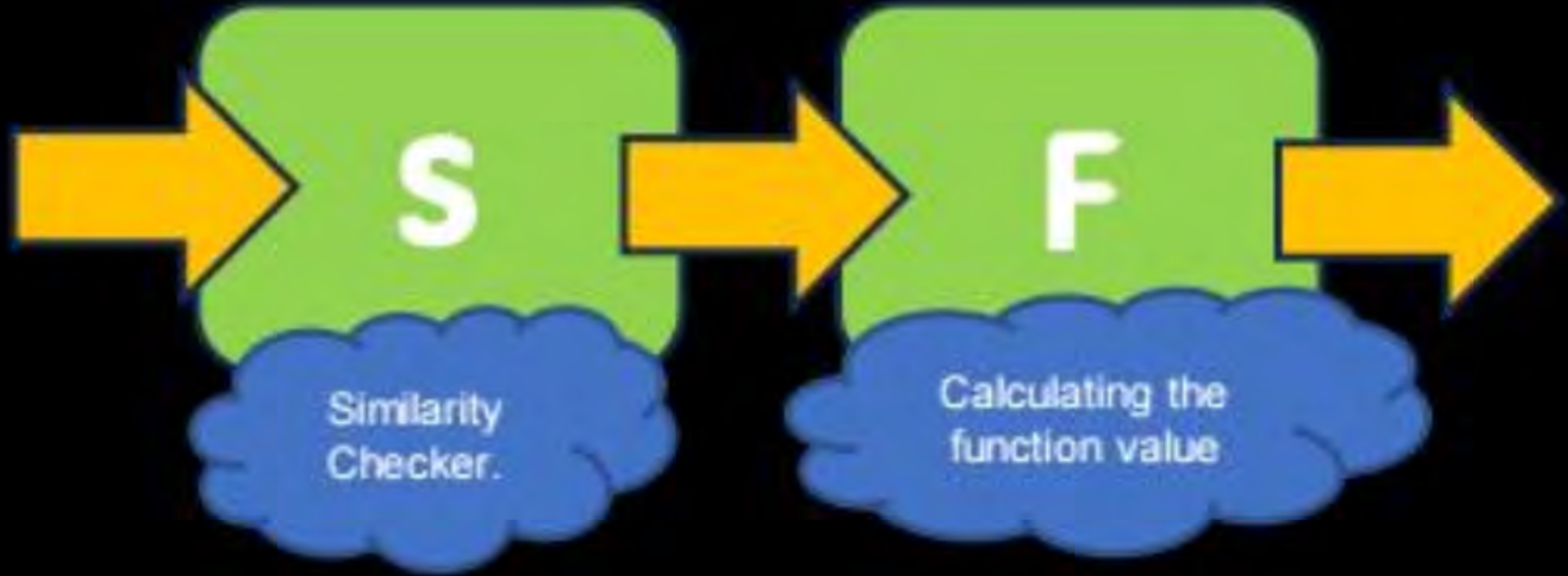
This method can be used for
1. Regression
2. Classification

We have instance based learning.

Nearest Neighbour Method

Nearest Neighbour method of Supervised Learning

But in Nearest Neighbour the process is as follows...



This is the lazy learner

This is the instance-based learning

Nearest Neighbour Method

Nearest Neighbour method of Supervised Learning

But in Nearest Neighbour the process is as follows...

This is non parametric in nature

It is widely disposable in real-life scenarios since it is non-parametric, meaning it does not make any underlying assumptions about the distribution of data

Nearest Neighbour Method

Nearest Neighbour method of Supervised Learning

But in Nearest Neighbour the process is as follows...

This can work as classifier and regressor...

Nearest Neighbour Method

Nearest Neighbour method of Supervised Learning

Similarity between the two points...

Similarity measure the distance between the two data points

Euclidian distance

Nearest Neighbour Method

Step 1 : find the distance of new point from all the other points in the data

Step 2 : Find the distance closest to the point under query

Step 3 : Now Simply assign the value of Y of the closest point to the new point

Finish

Nearest Neighbour Method

Nearest Neighbour method of Supervised Learning

Similarity between the two points...

So we first of all we find the training instance which is closest to the new point

Nearest Neighbour Method

Nearest Neighbour method of Supervised Learning

Now how to find the label or value of new point ?

Now finding the closest or similar instance we then give the same label or value to the new data point

Nearest Neighbour Method

Nearest Neighbour method of Supervised Learning

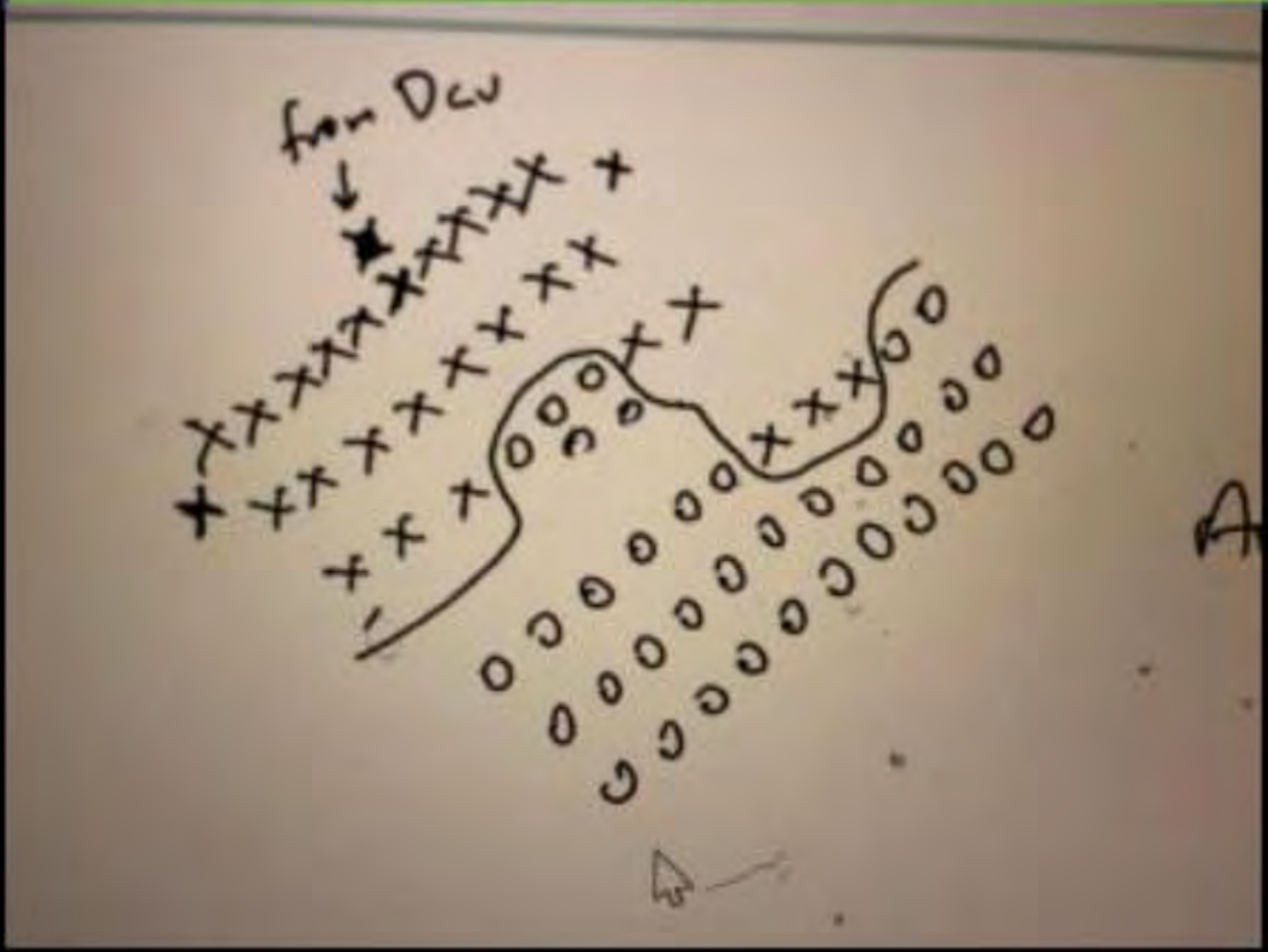
Decision Boundary in Classification Problem (NN)

The motive is to find out the region influenced by each point

Nearest Neighbour Method

Nearest Neighbour method of Supervised Learning

Thus we get this decision boundary (Voronoi Diagram)



So the final decision boundary will be ...

D:/Ace/Slides%20for%20session%201/knnalgo.html

DataSet D

x_1	x_2	y
1	7	0
1.5	8	0
2	9	0
2.2	7	0
1.8	10	0
2.3	8.5	0
2.9	11	0
6	7	1
6	8	1
6.5	6	1
7	6	1
7	7	1
8	8	1

$x \rightarrow y_i = 1$

$o \rightarrow y_i = 0$

What is impact of K on KNN algo.

More perfect

THANK - YOU