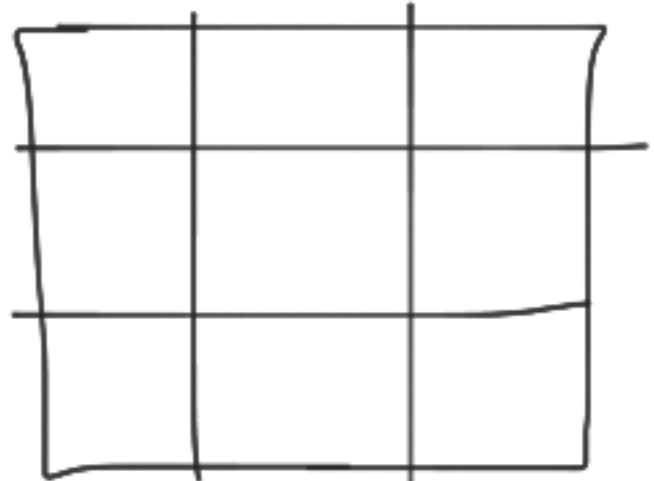
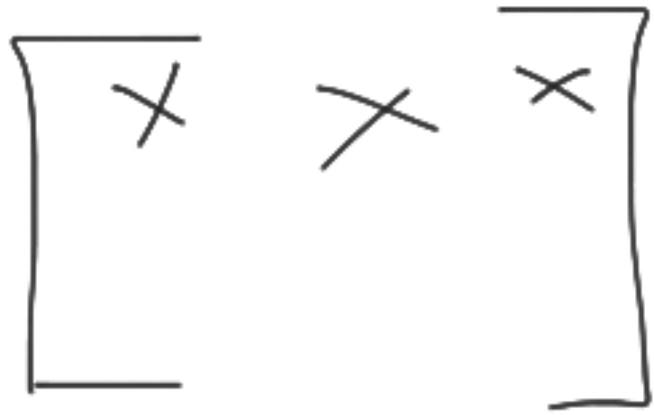


Logistics R. -> Sigmoid Func.  
Encoding -> Nominal & Ordinal

OHE & LE

Confusion Matrix



Dog  
Cat  
Horse



<u>n=165</u>	Predicted: NO	Predicted: YES
Actual: NO	<u>50</u>	Type I Error 10
	Type II Error 5	<u>100</u>

150/165



		Actual Class	
		Positive (P)	Negative (N)
Predicted Class	Positive (P)	✓ True Positive (TP)	False Positive (FP) <sup>Type I Error</sup>
	Negative (N)	<u>False Negative (FN)</u>	✓ True Negative (TN)

Type II Error

# Cross Validation (CV)

`train_test_split(X, y, test_size=0.2, random_state=42)`

80% -> Train  
20% -> Test

✓ 87%  
✓ 75%  
93%

Height	Weight	OB
		1
		1
		1
		1
		1
		1
		1
		0
		0
		0
		0

It is use to improve the cons of the train\_test\_split as it randomly selects the data (training & testing) which results in the fluctuation of the accuracy.

## Types of CV (Cross Validation)



### 1. LOOCV (LEAVE ONE OUT CV)

Nowadays, nobody use this type of CV

$n = 1000$

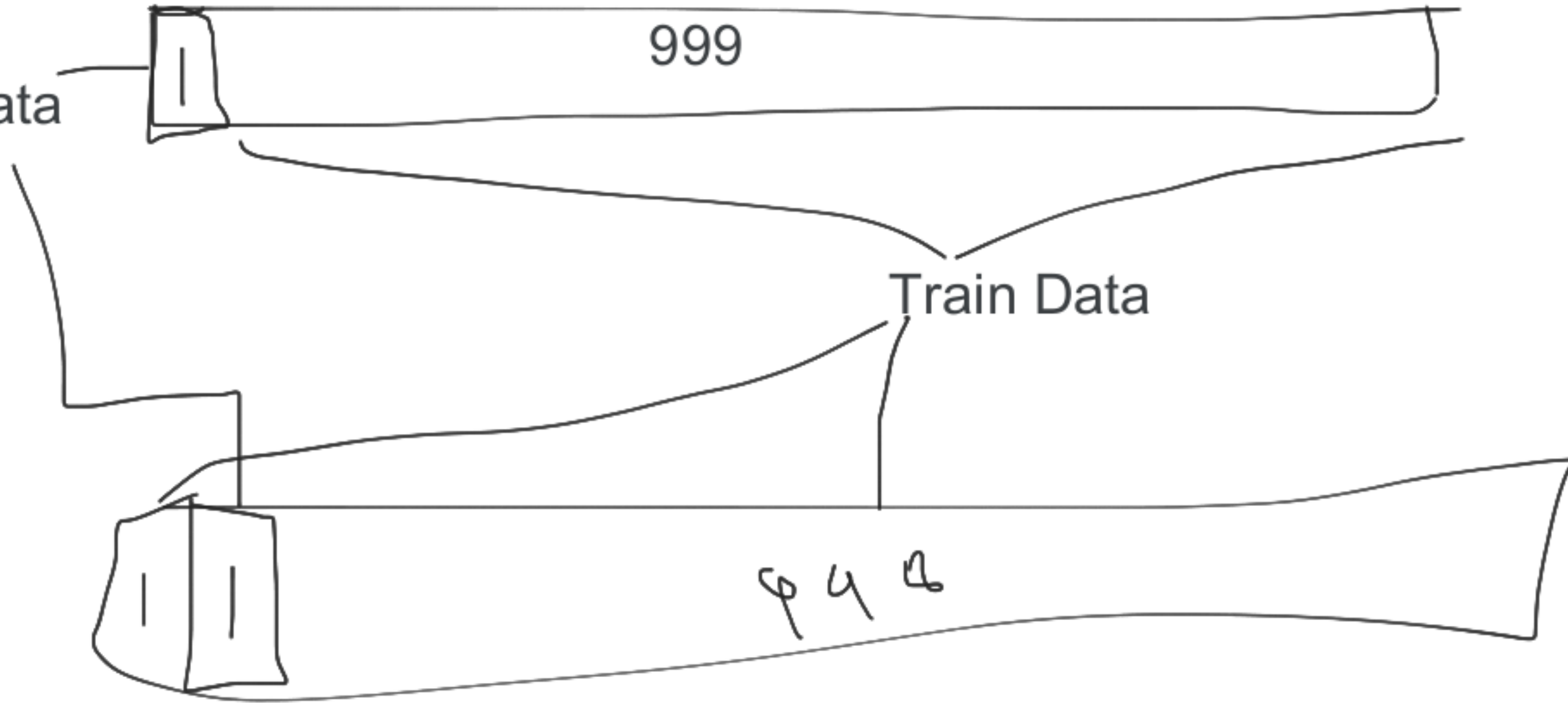
Test Data

999

Train Data

999

Many iteration is being used



2. K-FOLD CV

n = 1000

K = 5

84.34

85.89  
81.32

Train

Test  
1st Fold

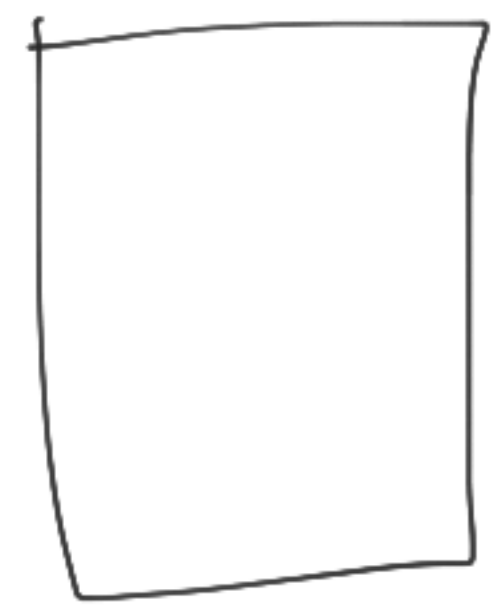
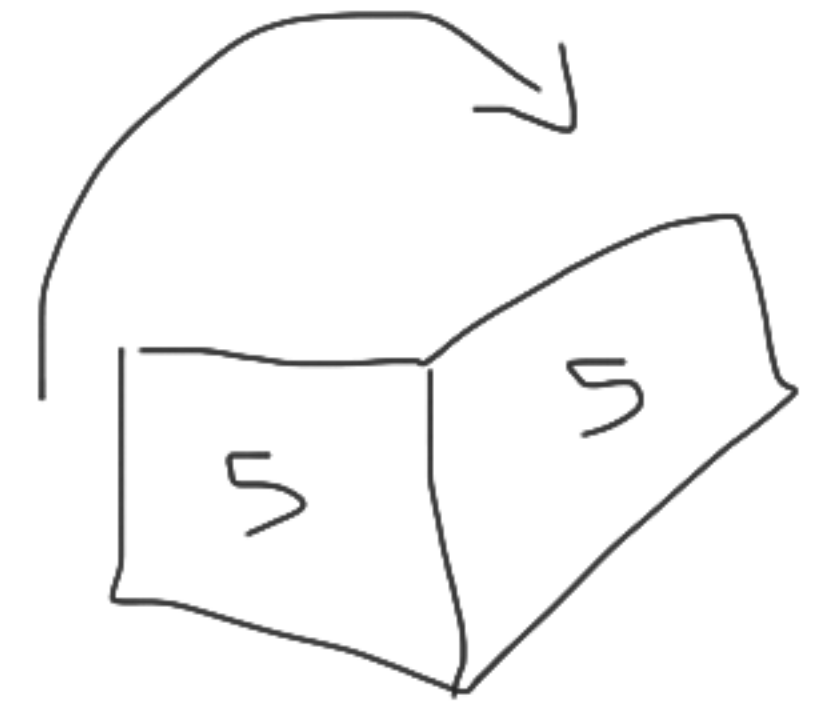
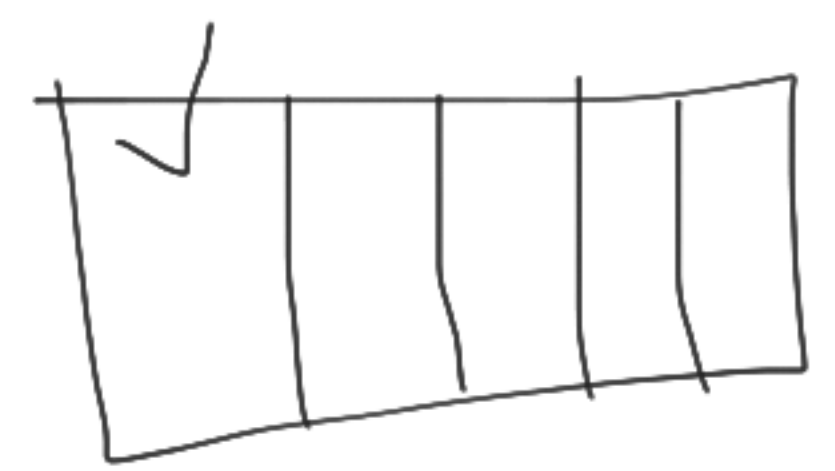


Test



Imbalanced Dataset

Train



### 3. Stratified K-FOLD CV $K = 5$

---

$n = 1000$

80

YES - 180 NO - 20	YES - NO -
----------------------	---------------

900 - YES

100 - NO

The main motive of the Stratified K-FOLD is to maintain the proportion of each classes in every training & testing dataset.

### 4. Time Series CV

$K = 3$  5 7

# K-Nearest Neighbour

✓ A - 3  
B - 4

X → (Math = 6, CS = 8) → P

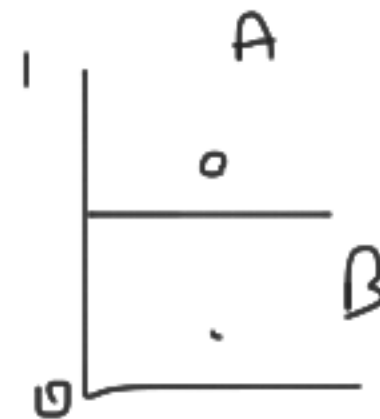
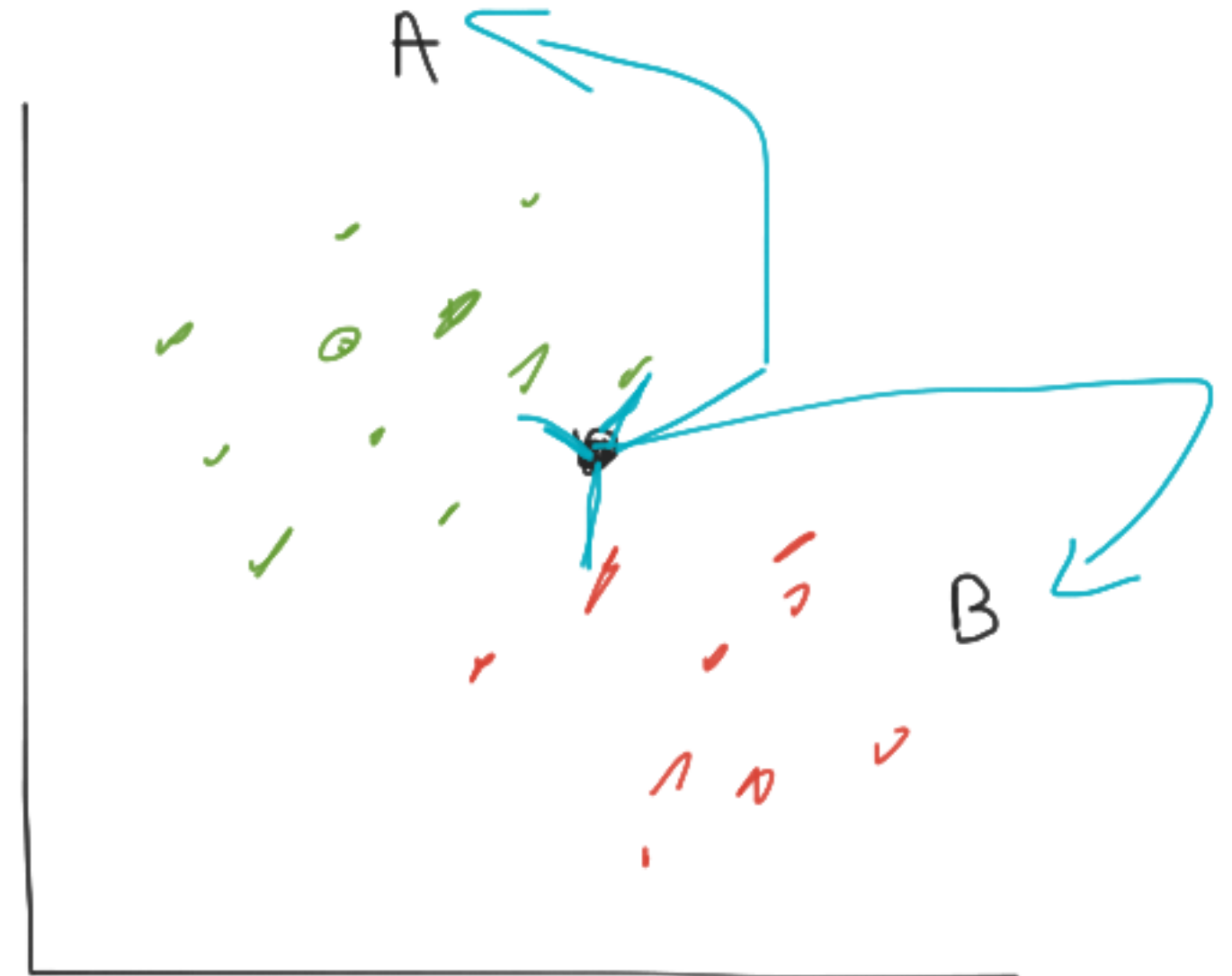
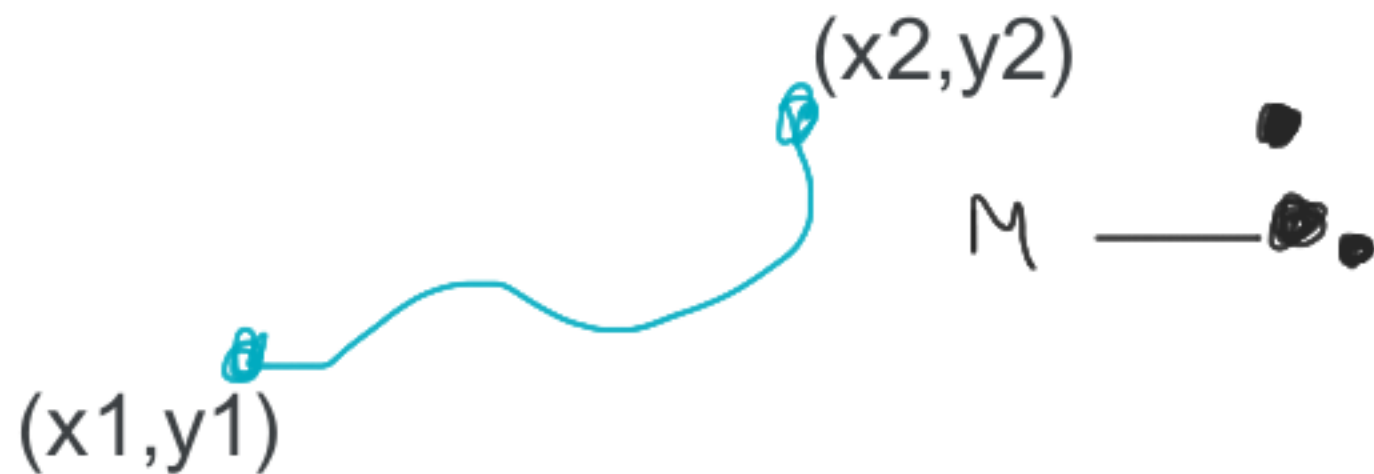
Maths CS Result ED

4	3	F	5.38
6	7	<del>P</del>	1✓
7	8	<del>P</del>	1✓
5	5	F	3.16
8	8	<del>P</del>	2✓

✓ P - 3  
F - 0

$$\sqrt{(6 - 4)^2 + (8 - 3)^2}$$

Euclidean D. =  $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$





How shall I choose the value of 'K' in KNN?

1. There is no methods or equations to find the value of 'K'. You have to use the hit & trail method to find out the value of 'K'.
2. You have to try to avoid choosing smaller values of 'K'
3. Always try to keep the values of 'K', as an odd number.

Pros:

1. Simple Algo
2. Gives better results
3. Multi classes cases

Cons:

1. You have to determine the value of K correctly
2. Storage will increase
3. High computation power is required
4. You should be aware about the distance finding algos

17 - KNN Pract  
18 - Naive - T & P  
19 - CLustering - KMEAN T&P  
20 - Association - APriori & ECLAT T &P  
21 - Intro of Deep

24 - ANN - T & P  
25 - CNN T  
26 - CNN P  
27 - RNN  
28 - NLP T & P

31 - GANs & AutoEncoders

Git & GitHub  
Resume Building  
LinkedIn Profiling  
Interview