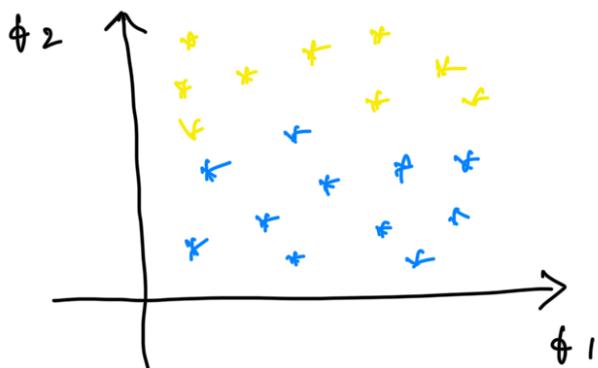


# K Nearest Neighbour (KNN)

① classification

② Regression

Classification



Binary category  
↓

$f_1$	$f_2$	$y$
-	-	0
-	-	1
-	-	0



Training Dataset

How do test the new data points?

① we have initialize the k value

- k can hold any value

$$k > 0 \dots \infty$$

$k \rightarrow$  represents number of neighbour.

this can be represented as hyperparameter.

② Find the  $k$  nearest neighbour for the test data.

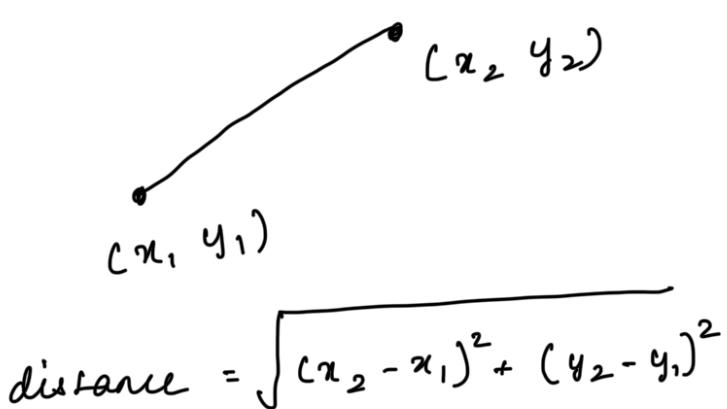
③ From those  $k$  values, how many neighbours belongs to which category.

How do we determine the  $k$  values?

This can be determined using distance formula.

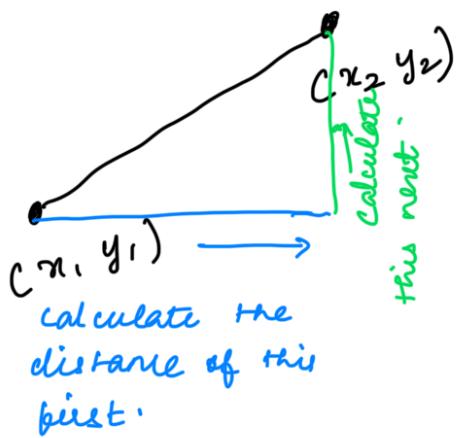
There are two types.

① Euclidian distance



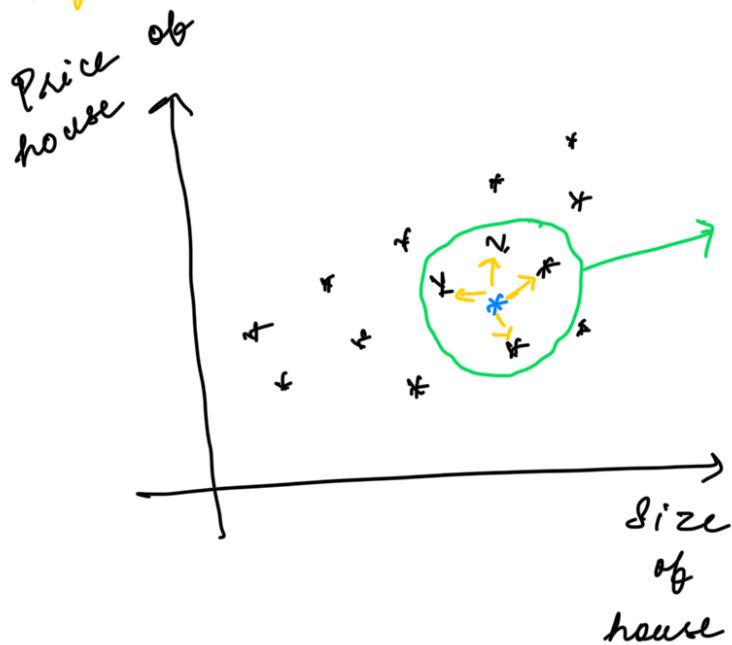
Basically, this is the distance between two coordinates.

② Manhattan Distance.



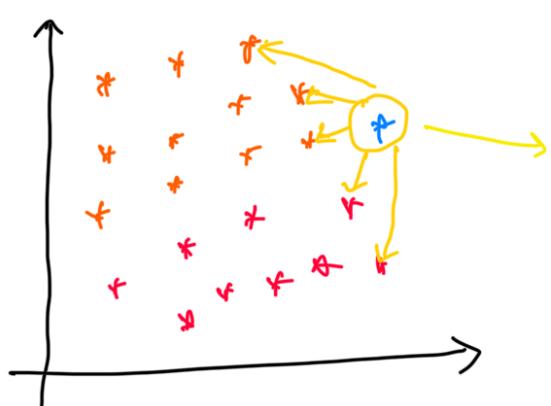
By calculating, the distance between blue and green, we will calculate the distance between two coordinates.

## Regression



lets say  $k = 5$ , when we have a new test data, we will calculate the distance of all the neighbour, and average of all those points to find out the output.

If there are outliers, then we will have to calculate median.



## Time complexity

$O(n)$   $\Rightarrow$  millions of data

- ① K D Tree
  - ② Ball Tree
- } Optimize

$\Downarrow$   
Binary tree.