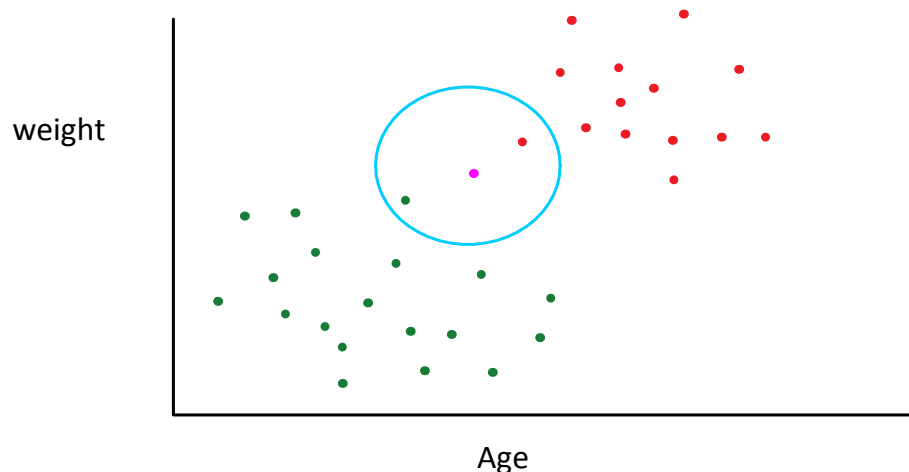


# K-Nearest neighbourhood

13 April 2022 11:55



If I give  $K = 1$ , it will start understanding which is the first nearest data point  
Then it will be predicted Red color--> Diabetic

If I give  $K = 2$ , it will start understanding which is the first two nearest data points

case a: If both are the points belongs to same category, it will predict Diabetic only

case b: If both are the points belongs to different category, it will predict only one thing  
it will be random based

If I give  $K = 3$ , it will start understanding which is the first three nearest data points

case a: all the three belongs to same category, then it will predict same

case b: 2 points say yes, 1 points say No --> Yes

2 points say No, 1 points say Yes --> No

Depends up on the number of samples those many k points are allowed.

What k values are suggestable

1.  $K = 3$  to 20
2. always try to choose k values as odd number

---

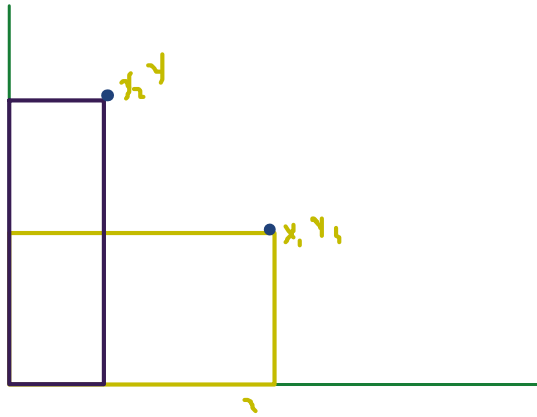
## NOTE:

If samples are in the data in classes are not equally proportioned then your ML can be learned with biased data only.

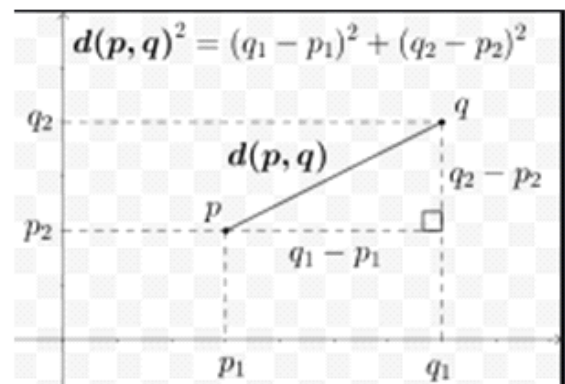
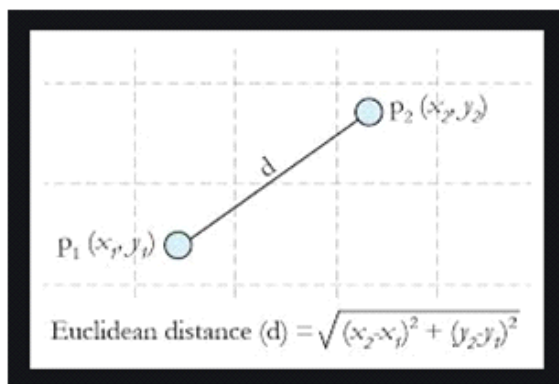
We have to make sure that our classes either 2 categories or  $\geq 2$  categories we have to ensure that those are to be balanced.

How to make it?

1. We have to collect more data so that classes are properly balanced.
  2. When there is no chance to collect some more data, we have to apply up sampling or down sampling methods before you apply machine learning model.
- 

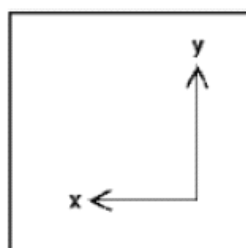


Euclidian distance

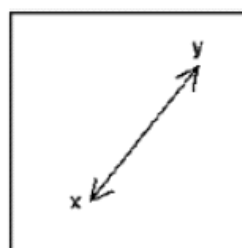


Manhattan distance formula

$$\text{Mdist} = |x_2 - x_1| + |y_2 - y_1|$$



Manhattan



Euclidean

Note: whenever you apply any distance method procedures make sure that your data should be

standardized.

D	E	F	G	H	I	J	K
				341.4	227.6		
				60	40		
				train	test	train	test
b	357	0.62742		211.668	141.112	0.62626	0.62626
m	212	0.37258		126.318	84.212	0.37374	0.37374
tota	569			337.986	225.324		

Screen clipping taken: 13-04-2022 12:59