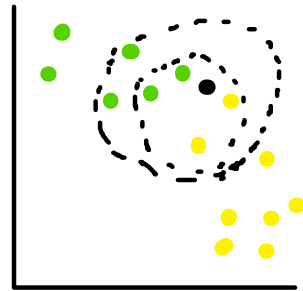Yellow = class A
green = class B

k = 5

class A = 4 →
class B = 1

for k = 5 = class A = 3
B = 2

k = 7 = class A = 3
B = 4

KNN is based on feature similarity.

To chose value of k

1) Total datapoints $(n) \Rightarrow$ sqrt(n) $= 3^6$

2) Odd value of k

When to use KNN

1) labelled data
2) data is noise free
3) data is small

Wt	Ht	Target BMI	
30	170	UW } — Class	
35	174	GV }	
n	176	40-176 }	noise
25	162	25-14 }	
21	168	81-108 }	

noise

lazy learner >> KNN

It doesnt learn from training dataset

It doesnt learn 0

$df1 = \dfrac{120, 6}{} = \dfrac{730}{} = 64 kb$
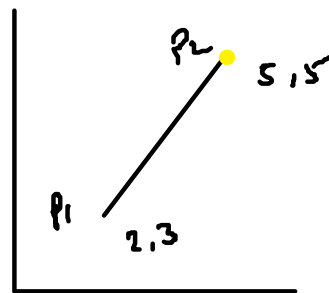
$df2 = \dfrac{120, 48}{} = \dfrac{4870}{} = 50 kb$

50000 (25)

20000

20, 15

P2 5,5
P1 2,3

Distance metrics
1) Euclidean distance
2) Manhattan distance
3) Minkowski
4) hamming
5) Jaccard
6) cosine
7) mahalobis

Minkowski

$$\left(\varepsilon \, |(x_i - y_i)|^p \right)^{1/p} \longrightarrow \text{deriving these distances}$$

$p=1 : |(x_i - y_i)| \Rightarrow$ Manhattan distance

$p=2 : \sqrt{(x_i - y_i)^2} \Rightarrow$ Euclidean distance
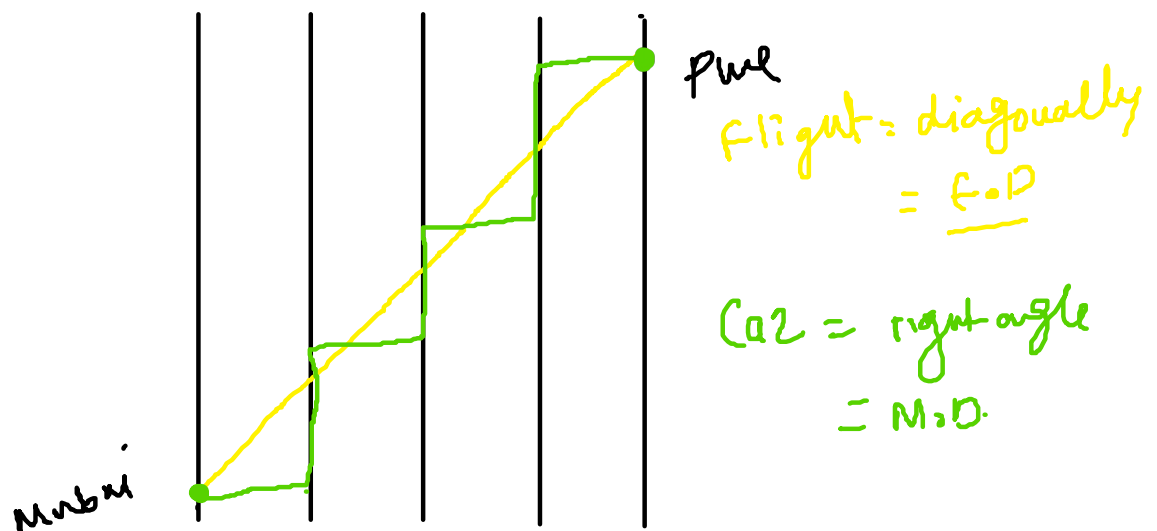
$E.D = \sqrt{\varepsilon \, (x_i - y_i)^2} \quad (p=2)$

$$E.D = \sqrt{\varepsilon \, (x_i - y_i)} \qquad \left(\underline{P = 2}\right)$$

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

$$= \sqrt{(2-5)^2 + (3-5)^2}$$

$$= \sqrt{9 + 4} = \sqrt{13} = \boxed{3.6}$$

$$MD = |(x_1 - x_2)'| + |y_1 - y_2|$$

$$= |2 - 5| + |3 - 5|$$

$$= 3 + 2$$

$$= 5$$

MD.   5,3

2,3

E.D. = Diagonally

Mumbai

Pune
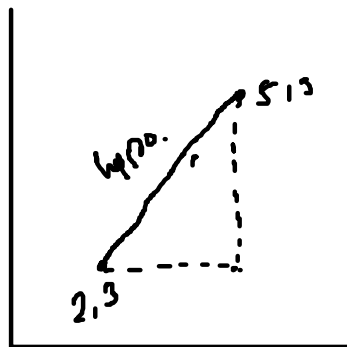
Flight = diagonally
= E.D

Car = right angle
= M.D.

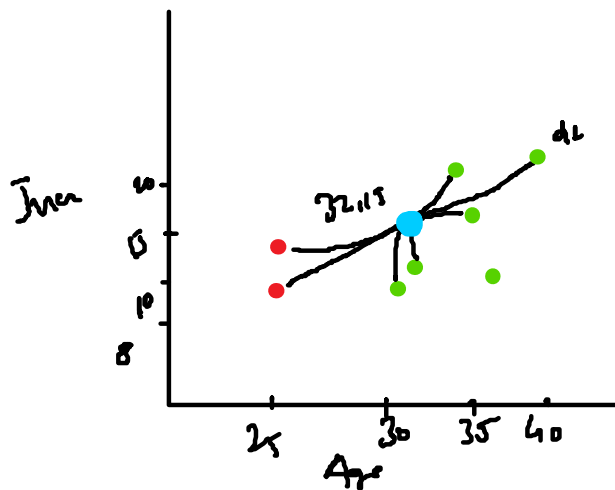|	Age	income (lacs)		Loan	6p.
d_1	40	20	A	9.3	
d_2	30	16	D	22	
d_3	25	68	A	49	
d_4	41	25	A	10	
d_5	26	09	D	85	
d_6	35	14	A	3.9	

Age — incm — $\underline{\text{Loan}}$

32 = 15 = ?

$k = 3 = d_2, d_3, d_6$

$d_1 = \sqrt{(32-40)^2 + (15-20)^2} = 9.3$

$d_2 = \sqrt{(32-30)^2 + (15-16)^2} = 2.2$

$d_3 = \sqrt{(32-25)^2 + (15-8)^2} = 9.5$

Industry appli:

1) Recommender system (Amazon, Youtube, Instagram)

2) Concept Search

Adhar

1234 5678 1249

PAN

hiii