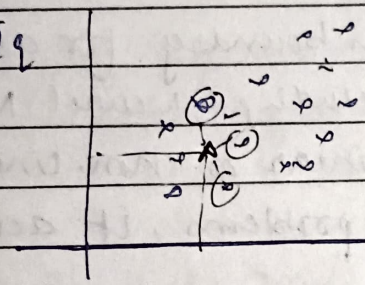
Page No.
Date

Advance KNN :-

→ KNN can be applied in both classification and regression.

• KNN Regressor :-

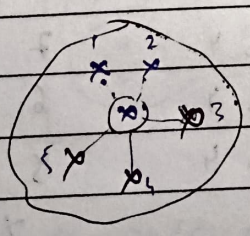
Suppose | Cgpa | Iq | Package |
↳ given & we have to find package



i) find dist. of query pt to all the pts.
ii) Sort them
iii) Let K = 3
& so take closest 3 pts.
→ take their package and find avg.

↳ In this if -
→ K value is large then underfitting
→ & K " " small " overfitting

→ weighted KNN :-



in uniform KNN we say it as Black as # black is more
But
In weighted KNN we calculate weight :-

$$\text{weight} = \frac{1}{\text{dist}}$$

	Dist	wt
1 - Blue	0.2	5
2 - Blue	0.5	2
3 - Black	1	1
4 - "	2	0.5
5 - "	3	0.33

wt of Blue = 5+2 = 7
wt of Black = 1.83 So it should be Blue

- Types of distances -

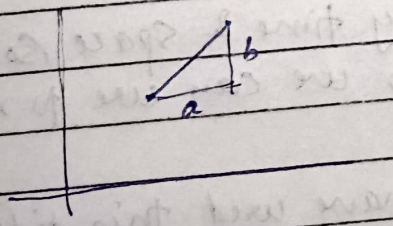
1) Euclidean dist -

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

$$\text{general (n-dim)} = \left[\sum_{i=1}^{n} (x_{2i} - x_{1i})^2\right]^{1/2}$$

↳ also called L2 Norm

2) Manhattan distance (taxi cab dist)



manhattan dist = a + b

$$m = \sum_{i=1}^{d} |x_{2i} - x_{1i}|$$

- Problem with euclidean dist
  - i) axis should be of same scale
  - ii) Curse of dimensionality

Page No.
Date

- In hyperparameter of KNN we have a, Parameter metric.
  default value of metric is minkowski → we have p = 2 (default)

→ p = 2 → euclidean dist (L2 norm)
  p = 1 → manhattan " (L1 norm)

$$\text{minkowski} = \left(\sum_{i=1}^{d} \left(|x_{2i} - x_{1i}| \right)^p \right)^{1/p}$$

p = 2.5 → L2.5 norm etc

- Time complexity →

  O(nd) n → # rows in training data
  d → " features " "

- Space complexity -
  O(nd)

→ So KNN takes very time & space so there are diff algo we can use to red it -

i) brute :- we have used this till n
ii) ball_tree
iii) kd_tree
  ↳ time complexity = O(d log n)

→ Normally we store data in arra but in kd_tree we store data in k-dimensional BST.