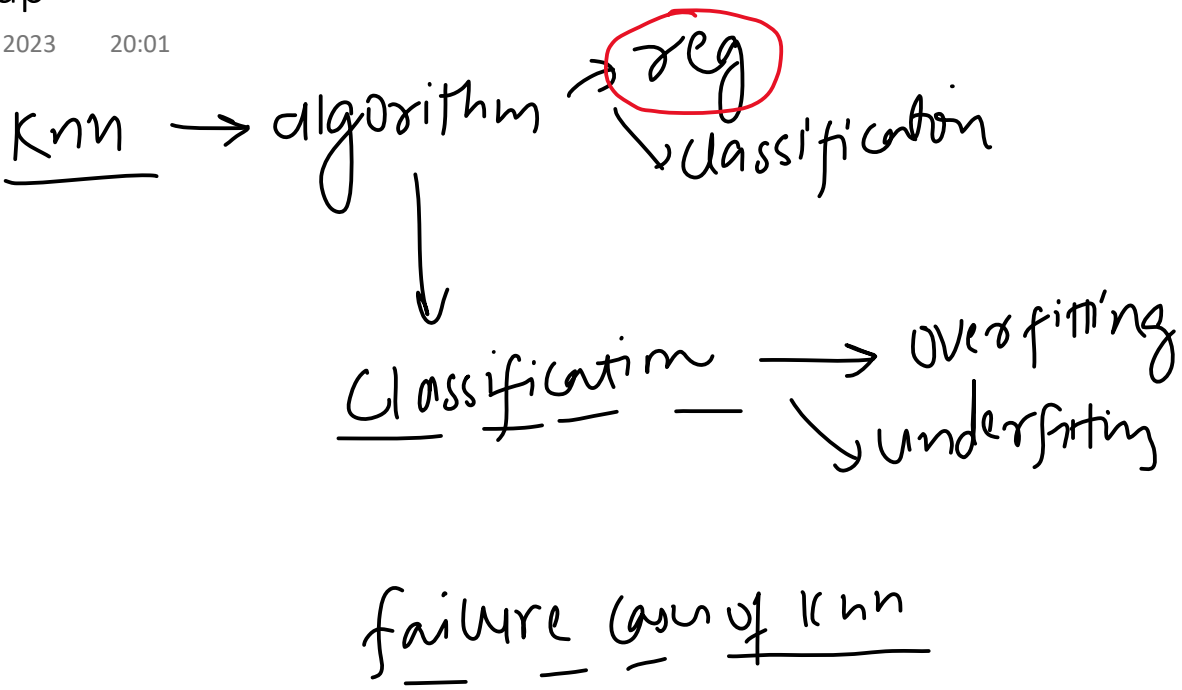
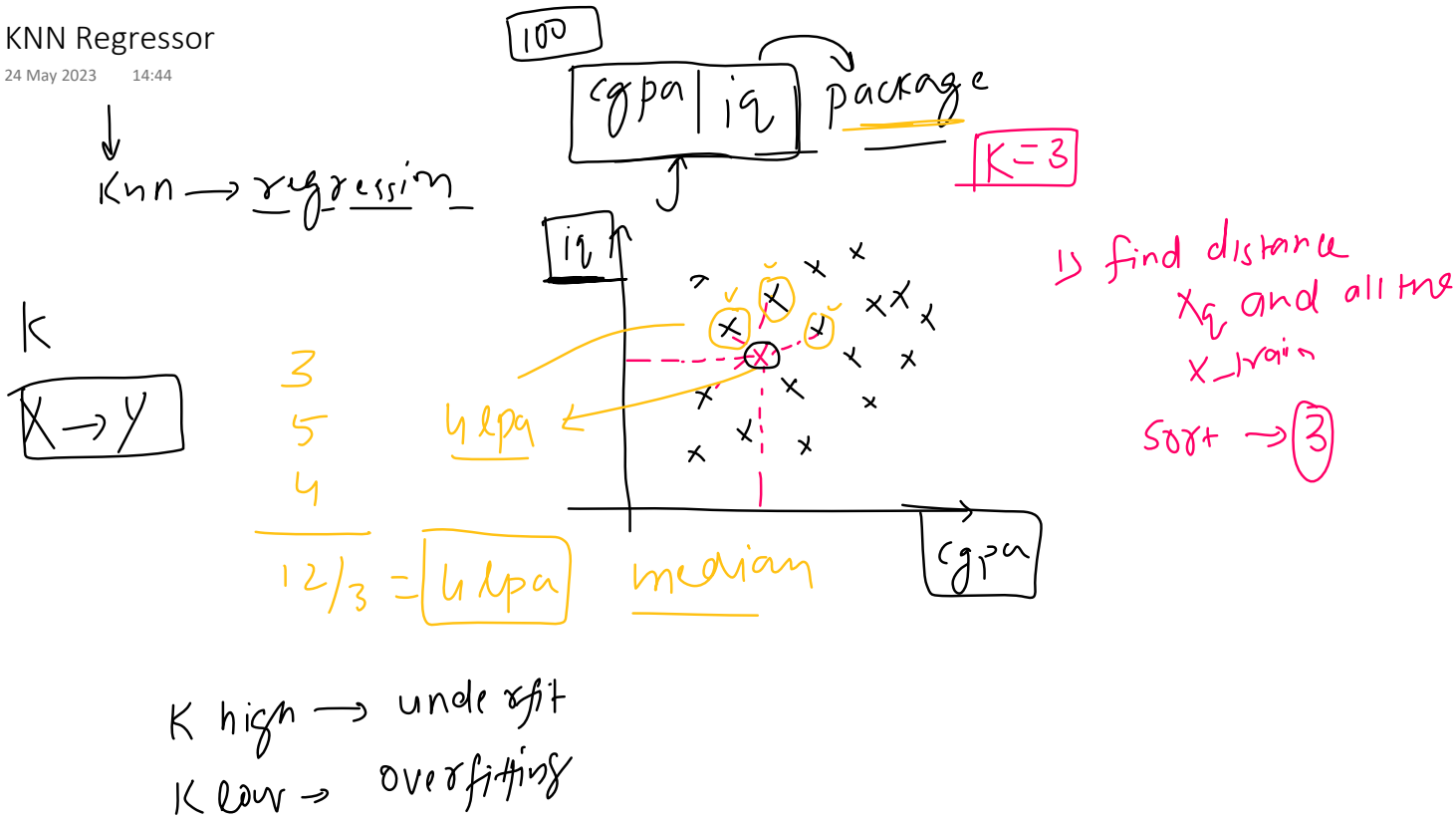
Recap

24 May 2023 20:01



KNN Regressor

24 May 2023 14:44

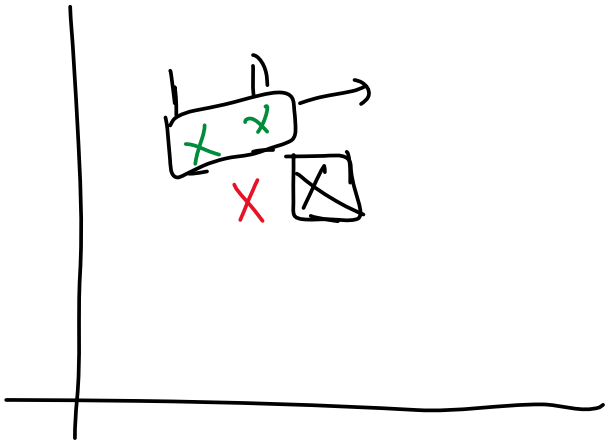


K high → underfit

K low → overfitting

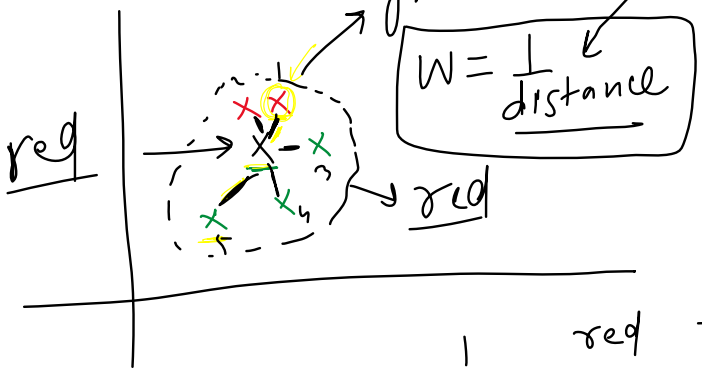
Hyperparameters

24 May 2023 14:44



Weighted KNN

24 May 2023 14:43



K=5 classification

$W = \frac{1}{distance}$

KNN → uniform

g → 2

g → 3 →

label → green

		dist	wt
1	red	0.2	5
2	red	0.5	2
3	green	1	1
4	green	2	0.5
5	green	3	0.33

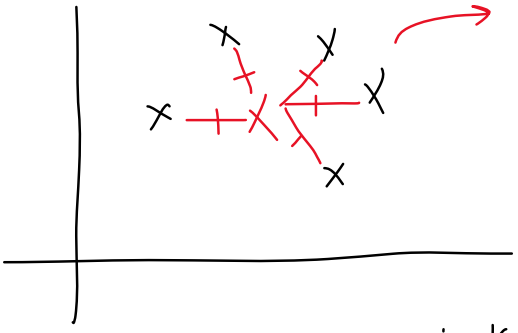
red → 5+2 = 7

green → 1 + 0.5 + 0.33

1.83

Types of Distances

24 May 2023 14:44

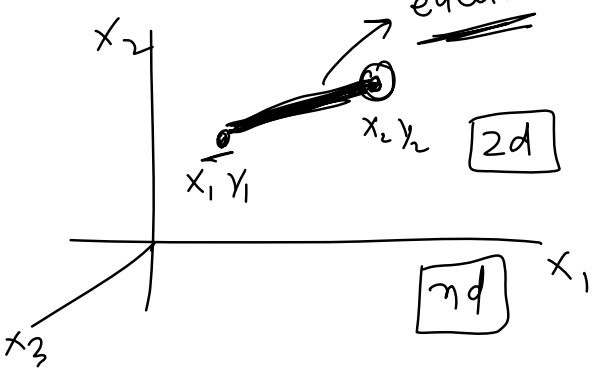


euclidean → types of distance

minkowski → eucli L2, man L1

metric → minkowski → p → 2 → euclidean

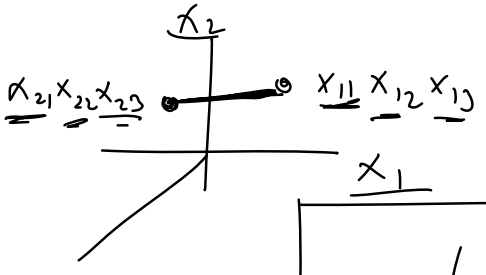
Distances



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

$dist = \sqrt{\sum_{i=1}^{d}(x_{2i}-x_{1i})^2}$ → 3 dim x_1 x_2 x_3

$x_1, x_2, x_3 \ldots x_d$



$\sqrt{(x_{21}-x_{11})^2+(x_{22}-x_{12})^2+(x_{23}-x_{13})^2}$

$$dist = \left(\sum_{i=1}^{d}(x_{2i}-x_{1i})^2\right)^{1/2}$$

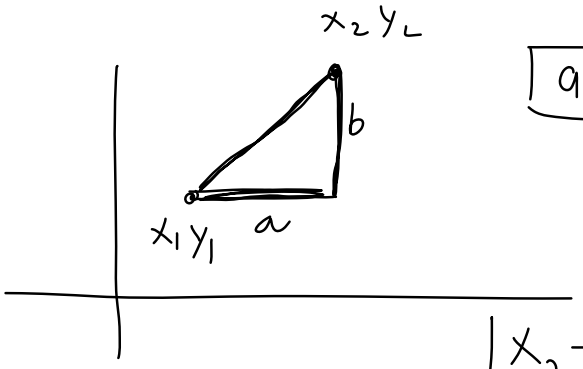
n dim points

Vectors norm → L2 norm

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

Manhattan distance

Manhattan distance

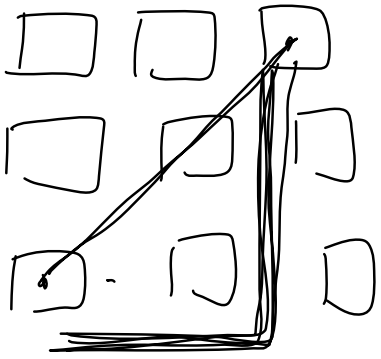


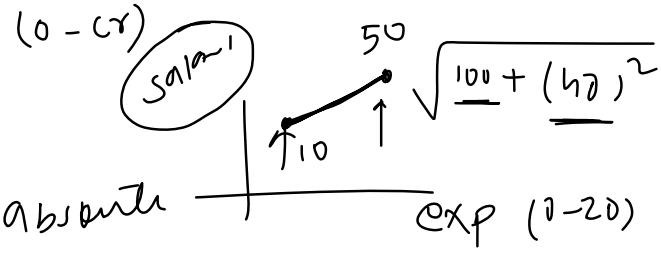
a+b → manhattan distance

↓

taxi cab distance

$$|x_2 - x_1| + |y_2 - y_1| \rightarrow \text{manh}$$





problems → euclidean

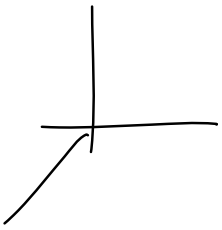
1) same scale ←

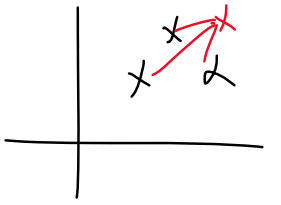
2) curse of dimensionality

d = 100

euclidean reliable

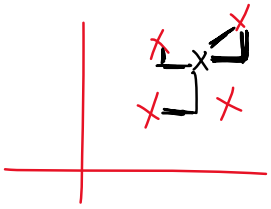
↑ manhattan





p = 2 → euclidean

p = 1 manhattan



minkowski

$$\left(\sum_{i=1}^{d} (x_{2i} - x_{1i})^2 \right)^{1/2} \qquad p = 2 \quad L2$$

euclj →

$$\left(\sum_{i=1}^{d} |x_{2i} - x_{1i}|^1 \right)^{1/1} \qquad p = 1 \quad L1$$

man

man

$$\left(\sum_{i=1} |X_{2i} - X_{1i}| \right)$$

general

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

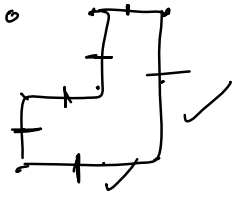
0 → higher

$p > 0$

← 2 (each
← 1 (man

3 L3

2.5 L2.5



Space and Time Complexity

24 May 2023 14:46

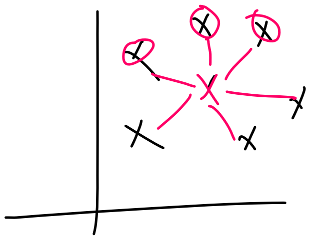
Knn → slow algorith

↓

product → train

↑

production



Sort ↓ K nearst ↓ major ↓ ← result

→ 2 min

time complexity

outline $f(n)$ dons

n → # no of rows

Time complexity

$O(nd)$

n → # of rows in training data

d → # features

$O(nd)$

1 cr

$x_1 \; x_2 \; x_3 \; \ldots \; x_d \mid y$

(10L, 1000)

2×10^8 → distance

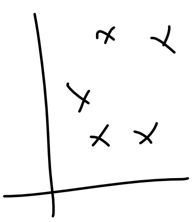
euclid

$$\sqrt{\sum_{i=1}^{d} (x_{2i} - x_{1i})^2}$$

d times (2) → 1 cr

space compexity

1GB → ram



→ $O(nd)$

KD-Tree

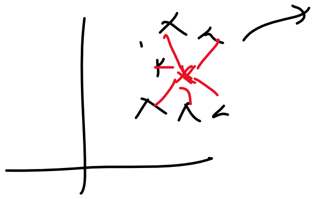
24 May 2023 14:44

time complexity → O(d log n)

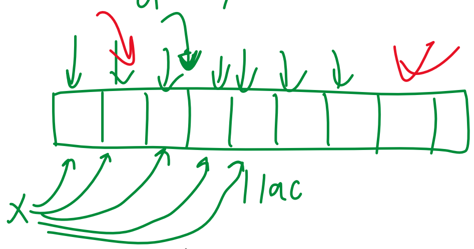
d → cols, log n

n → log n

train → 1 lac



array → data structure

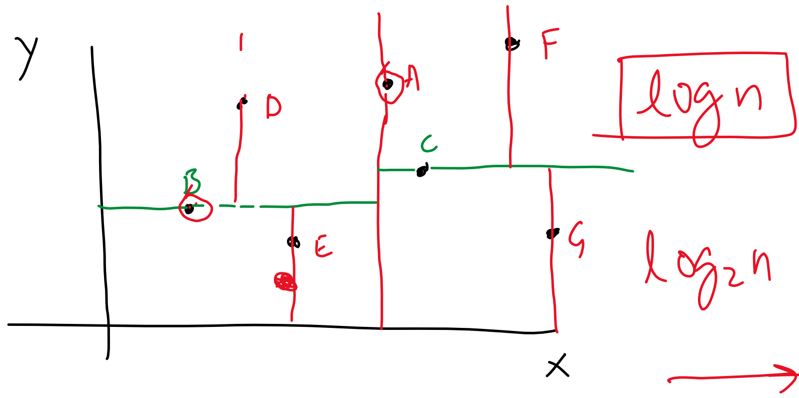


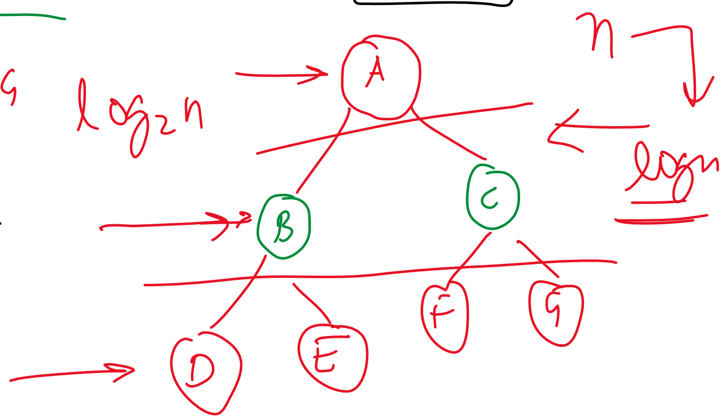
Binary search tree

k-d tree

K-d tree → x or y

2d-tree





xq → (5, 10)