8:11

Session on KNN Part 1 HOME INSERT DRAW VIEW

Stop

KNN Intuition

22 May 2023 19:14

You are the average of the five people you spend the most time with

- Jim Rohn

Knn → simple/elegant

100 student

cgpa	iq	placement
8	80	1
7	70	0
.	.	.
.	.	.
.	.	.

ml model

{ cgpa, iq } → placement

knn

8:15

Session on KNN Part 1 HOME INSERT DRAW VIEW

Stop

- Jim Rohn

euclidean

knn

iq

cgpa

class label 1

x → 1

x → 0

K=3

100 distance → 3 nearest neighbors

↳ sort ascending

closest

↳ majority count

1 1 0 → (1)

8:16

Session on KNN Part 1 HOME INSERT DRAW VIEW

Stop

2D → n-D

cgpa

closest

↳ majority count

1 1 0 → (1)

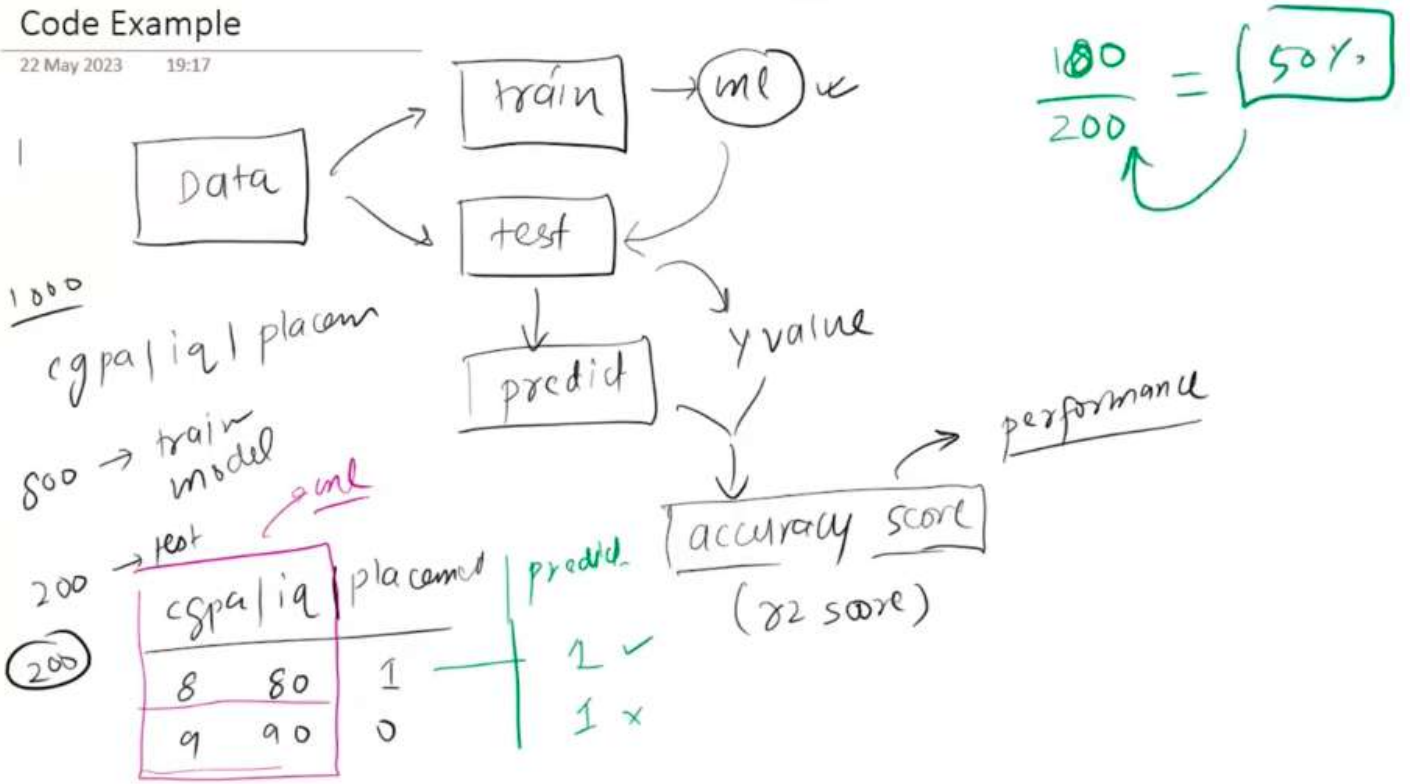
↑ ↑ ↑

$x_1 \mid x_2 \mid x_3 \ldots x_n$ n-dim

3 vectors → majority count → class label of query point

Code Example

22 May 2023 19:17



OneNote for Windows 10

Nitish Singh

Home Insert Draw View Help

Share

Calibri 11 B I U

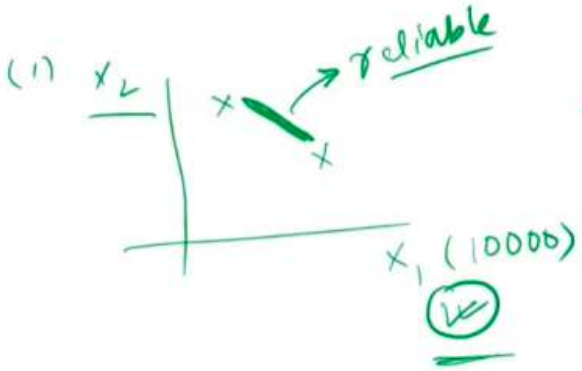
Heading 1

DSMP 2022-23

Session 3...
Session 4...
Session 1...
Session on...
Session 2...
Session 3...
Session on...
Session on...
Doubt Cle...
New Secti...
Session on...
Session on...
Session on...
Session on...
Session on...
Session on...
Session on...
Session on...
Session on...
Session on...
Session on...

KNN Intuition
Code Example
How to select K?
Decision Surface
Overfitting and Underfit...
Weighted KNN
KNN for Regression
Limitations of KNN

Add sect...
Add page



same scale

age → M, σ

72 → 7

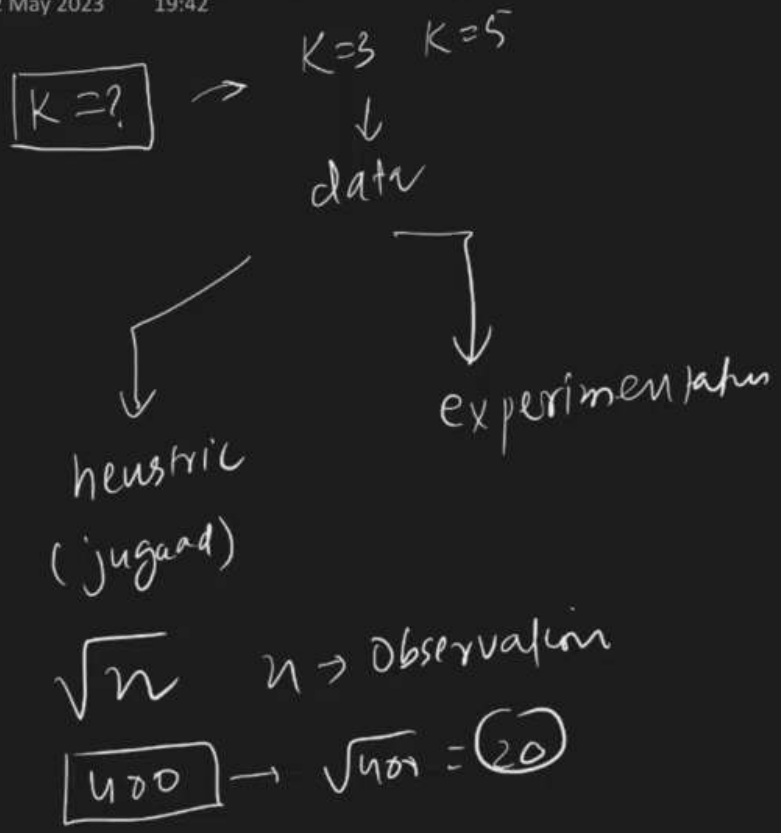
73

61

37

How to select K?

22 May 2023 19:42



9:08 23%

Session on KNN Part 1 HOME INSERT DRAW VIEW

Stop

data

heuristic (jugaad)

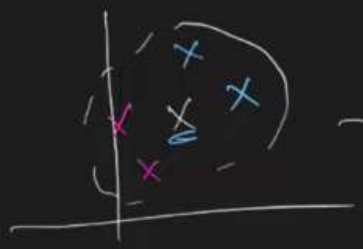
$\sqrt{n}$ $n \rightarrow$ Observation

400 $\rightarrow \sqrt{400} = 20$

even value for K

20 $\rightarrow$ 19 $\rightarrow$ 21

experimentation



K=4

2(1) 2(0)

conflict

Cross-validation

1000 = n

800 (train)

200 (test)

Stop

(jugaad)

$\sqrt{n}$ $n \rightarrow$ Observation

400 $\rightarrow$ $\sqrt{400} = 20$

even value for k

20 $\rightarrow$ 19 $\rightarrow$ 21

Cross-validation

1000 = n

800 (train) 200 (test)

run $\rightarrow$ accuracy

diff knn

knn $\rightarrow$ 1 $\rightarrow$ 83

knn $\rightarrow$ 2 $\rightarrow$ 81

knn $\rightarrow$ 3 $\rightarrow$ 87

knn $\rightarrow$ 25 $\rightarrow$ 91

10:44 38%

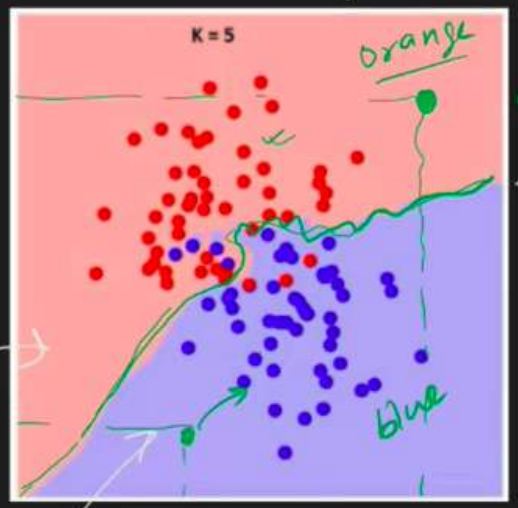
Session on KNN Part 1 HOME INSERT DRAW VIEW

Stop

[Decision Surface] → tool → classification → knn → svm, lr, dt, nn

22 May 2023 19:15

1D 2D 3D



coordinate

disc video

cgpa | iq | place binary [0,1]

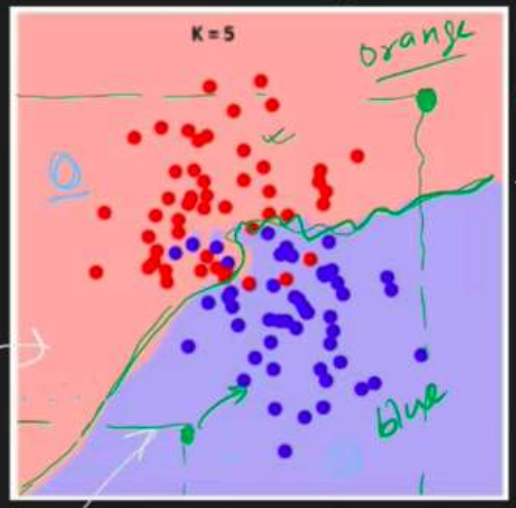
training points

cgpa

10:46 38%

Session on KNN Part 1 HOME INSERT DRAW VIEW

Stop



coordinate

disc video

cgpa | iq | place binary [0,1]

training points

cgpa

pixel on anim

generate a numpy meshgrid

knn train → 1
→ 0

orange

blue

7:53 49%

Session on KNN Part 1 HOME INSERT DRAW VIEW

Stop

Overfitting and Underfitting in KNN

22 May 2023 19:15

cgpa | iq | placed

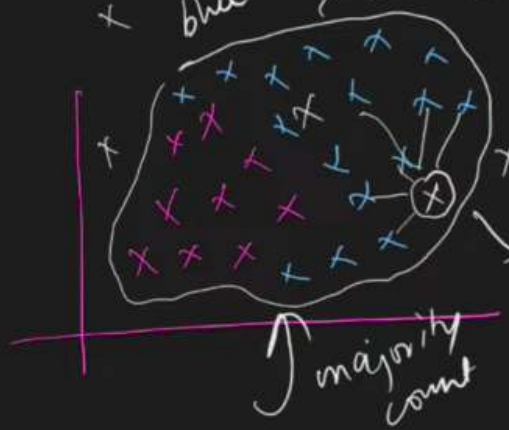
200 student

K=1

→ minor changes → overfitting

↓

high variance



blue

neighbour

K=? n=200

K=200

simplistic → blue

↓

underfitting

blue > pink

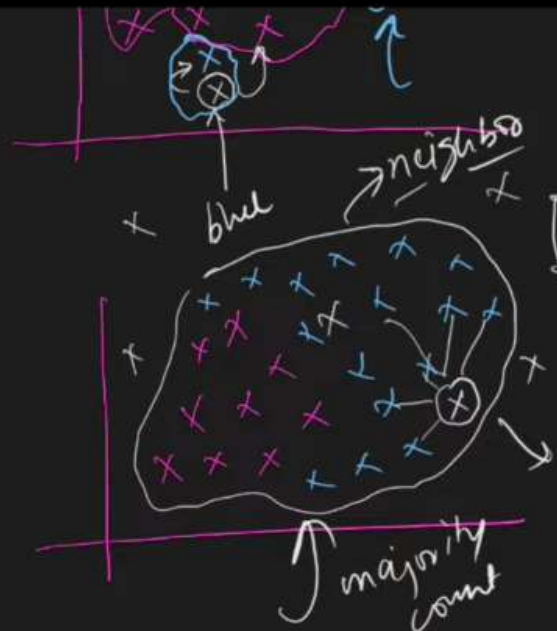
majority count

blue

7:53 49%

Session on KNN Part 1 HOME INSERT DRAW VIEW

Stop



K=?

n=200

K=200

simplistic → blue

underfitting

K = low

overfitting

K = high

underfitting

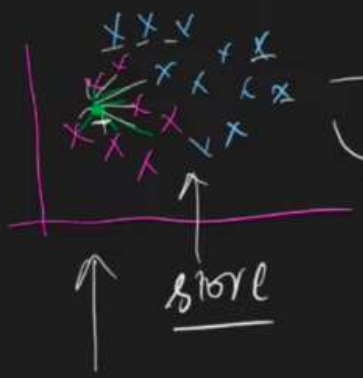
failure case

(500000, 100)

1) large datasets → n = 5L , f = 100

prediction ↓ slow → dataset

↳ Knn is a lazy learning techn



→ query point → prediction

↳ train → nothing

store

5L distance → sort → majority

low latency 3 sec → slows

8:40 46%

Session on KNN Part 1 HOME INSERT DRAW VIEW

Stop

store

Low latency [3 sec] → slows

2) High dim data → [f = 500] →

curse of dimensionality → distance concept → reliable

KNN

8:43 45%

Session on KNN Part 1 HOME INSERT DRAW VIEW

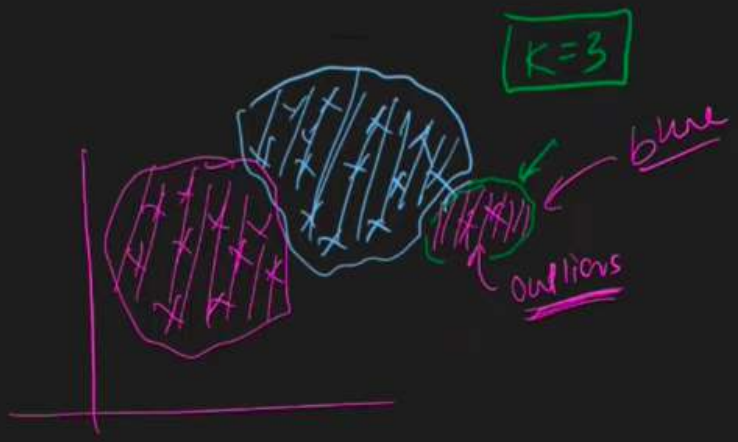
Stop

curse of dimensionality

concept

knn

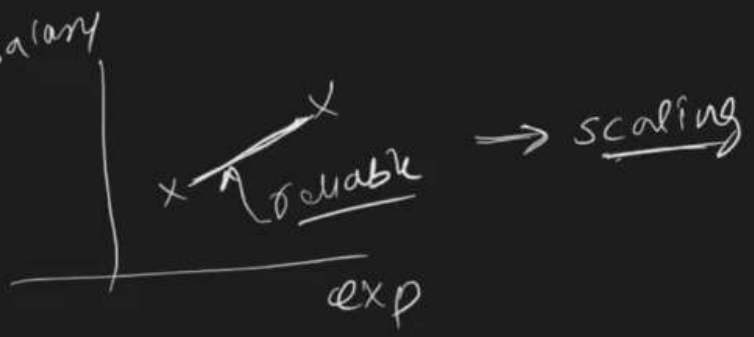
3) Outliers



4) Non- homogeneous scales

dominate

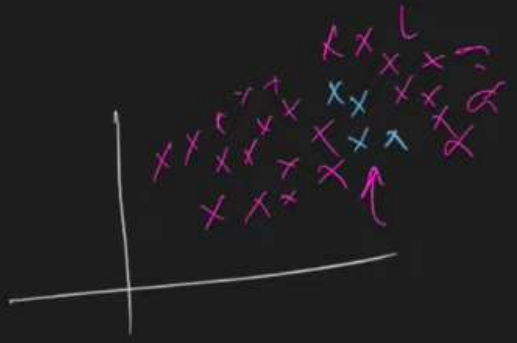
exp	salary	fire
0-25	20k-1cr	



→ scaling

0-25 20k-

5) Imbalanced dataset



↳ Yes → 98%

No → 2%

↳ biased → print

6) Inference and not for prediction

Stop

No → 2%

↳ biased → print

6) Inference and not for prediction