Dr/Hesham Essam

we have Two Types of Data
structured Data → Tables
un structured Data → image - Video - Audio

any one of Them can be labeled or unlabeled.

any Data Can be labeled or not but it is about your prespective

ساعات تبقى محتاج تعرف هل الشخص دة ممكن يسقط ولا لا ...
هل الـ Data كافية لأنك تجاوب على السؤال

هتعرض البيانات بتاعتك على استشاري، وهو لو عرف يجاوب بشكل كامل ← أمان ☺

Feature you use → get it From expert.

Semisupervised : will label small percentage of Data and predict other it

10 M → 100 K label
↳ 9.9 M predict

بنسبة كبيرة error بتاعة عالي.

Date / / No

Clustering:
blind classification
unlabeled Data.

3	150	4	15
3	160	5	12
4	155	4	10
4	150	5	15

5K

(720, 530, 115) (11, 170, 230) etc

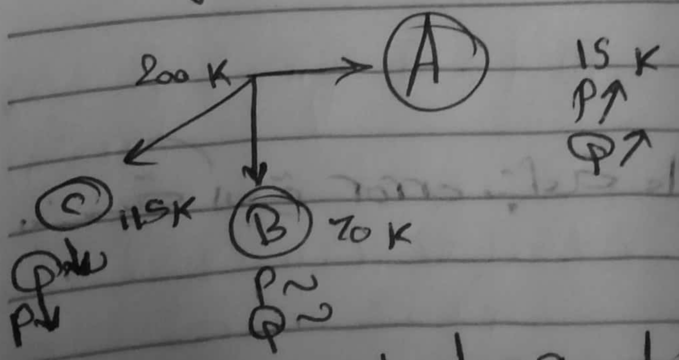
number of elements in every group different

لو طلعوا نفس العدد ← شك في نفسك

الفكرة من ال clustering بناءا على اي؟ هدف ← شاطر - طويل - دكر
(etc)
العسكرية

Many projects of ML → Medical - Business - Ministry

clustering will Make dealing with people easy



client segmentation — العدد ونفسيته ليه معنى

Date / / No

Body	Math	E
A1 1	A1 1	A1 2
B1 2	B1 1	B1 2
C1 1	C1 2	C1 2

4

number of clusters → predefined.

K-means clustering:

K → Number

exp

Salary age

Ai: لو جاءت بتريبها كدا عش

steps:

Random Centers (n of clusters)

calculate Distance with Centers to each Point.

labeling Points to Centers by Min D.

Moving center to mean of each Feature in group.

(Repeat until n of iter)

hint

Try to do clustering to Main Data to (9) ⇆ Compare

clustering to Sub Data to (9)

Date / / No

Evaluation

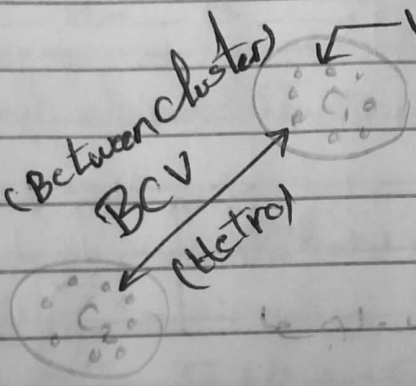
Still hard.

SSD &

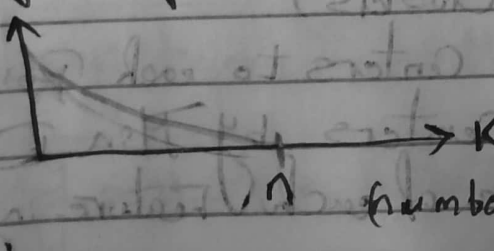
Summation squared Distance.

$SSD = \frac{1}{m} Distance(C_1, Points\ of\ group\ C_1) + \dots$

↳ Min (best)



So you try to Minimize (WCV) & Max (BCV)



*So you try to get (K) between (0 → n) but not (n) or (Zero)

* best (K) will be best by way → elbow

Draw graph