

Unsupervised learning:

$$y = f(x)$$

Supervised

x_1		y_1
x_2		y_2
x_3		y_3

Semi

class	label
Rain for past 4 days	Rain for tomorrow

Unsupervised → No class labels → Don't know what to do
for team.

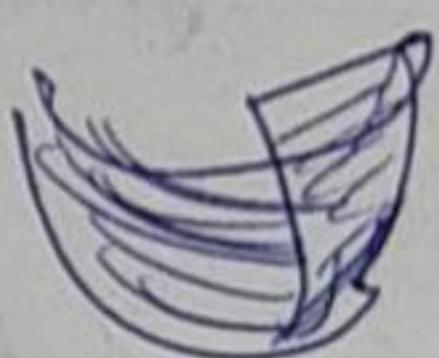
On

k-means clustering

Generic - prob: Don't know abt clustering

Now, given K , find the clusters

The ids of clusters don't matter



N - samples

K - clusters → $2^{n-K} = 2^{n-1}$

(2) → $A \leftrightarrow B$

If we could give a measure of clustering.

If we could give a brute force through 2^{n-1}

then we could

If we know the means

$|p - \mu_1|, |p - \mu_2|$

↓ ↓

find which is closer

Given means, clusters

can be identified

assign closer

given clusters, can i find label.

A → B (Means)
(Cluster)

& means

So, pick randomly & means
find the clusters

Data Clustering
Jain & Dubes