→ Class start 9:05 PM

ML 1·2 Unsupervised ML

→ Clustering (K Means, K means++, Hieanchial, DBSCAN, GMM)

→ Anomaly Detection like Isolation forest etc.

→ High Dimension Visualization (PCA, t-SNE, UMAP,

Unsupervised ML

→ Supervised ML → { Features
target or labels, or ground truth

→ Unsupervised ML → Features

→ No target or label

Classification

Binary

$$\left\{ (x_i, y_i)_{i=1}^{m}, \; x_i \in R^d, \; y_i \in (0,1) \right\}$$

Multi-class

$y_i \in S$ → Set of Class

Regression

$y_i \in R$

$$\{ x_i{}_{i=1}^{m} , \; x_i \in R^d \}$$

Examples of Unsupervised ML

→ Anamaly Detection / Fraud Detection

→ Clustering problem

→ Dimensionality Reduction like PCA

→ Recom. System like MF

→ Word-2-vec (NLP)

→ Autoencoder (CV)

Clustering

→ Process of grouping any kind of data based on similarity of these features

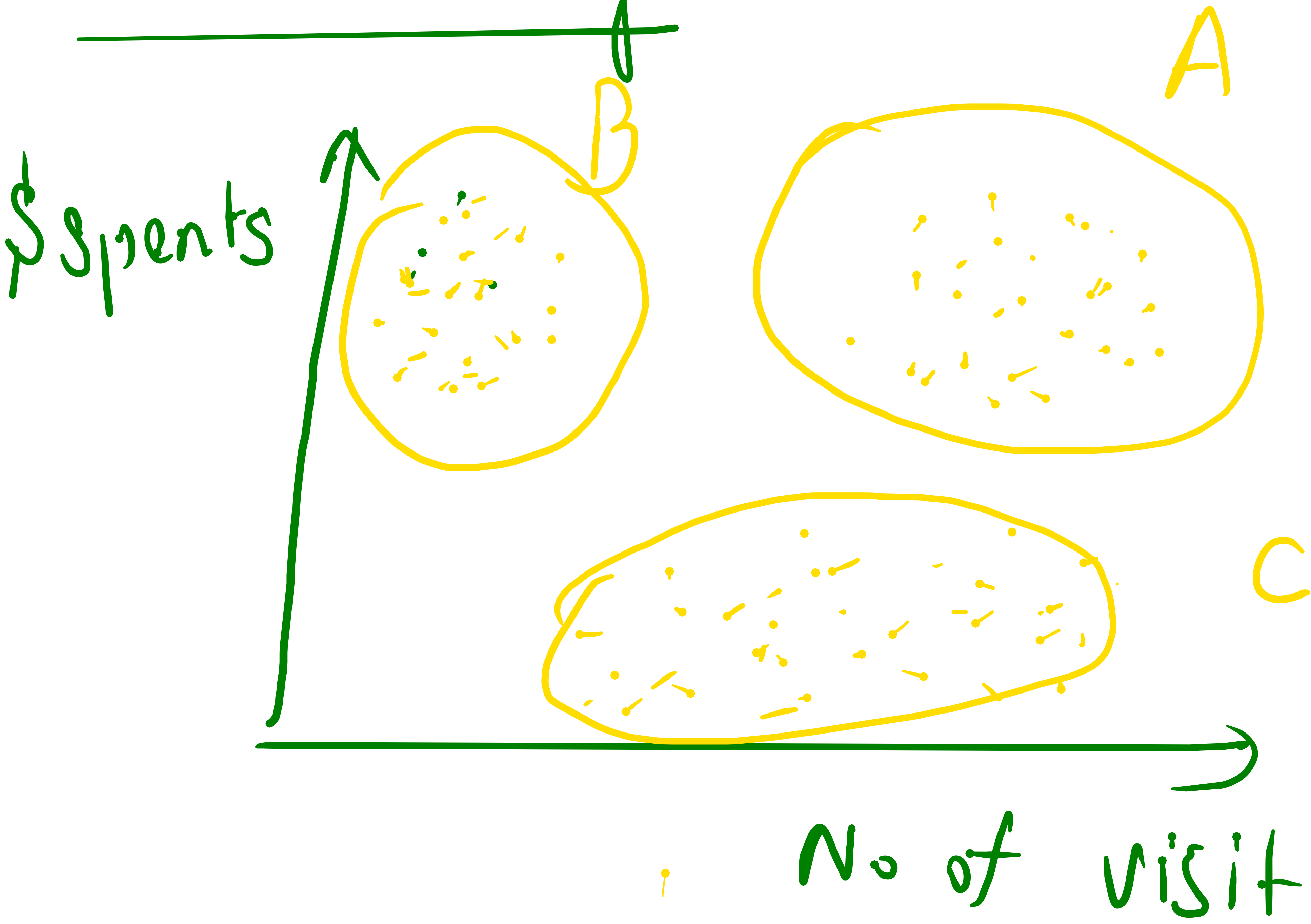
Eg

→ Customer Segmentation / Product Segmentation

→ Detecting similar stock

→ Google Photos groups similar in galaxy

Clustering



A : Heavy shopper, visit a lot, spent a lot

B : Rich people / Impulse buyer

C : Window shopper

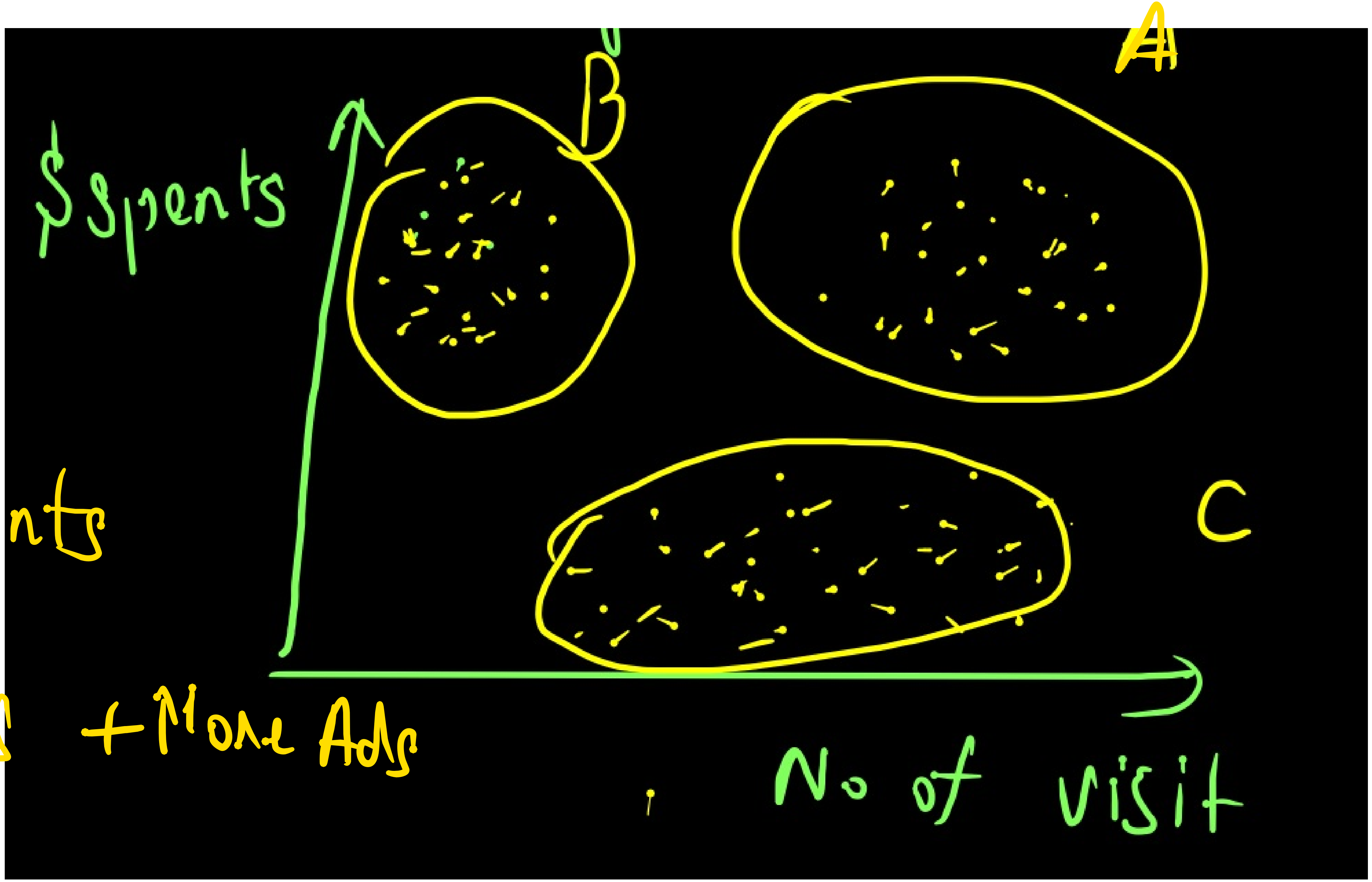
→ No of Discount coupon

→ No of Ads

B → More Ads

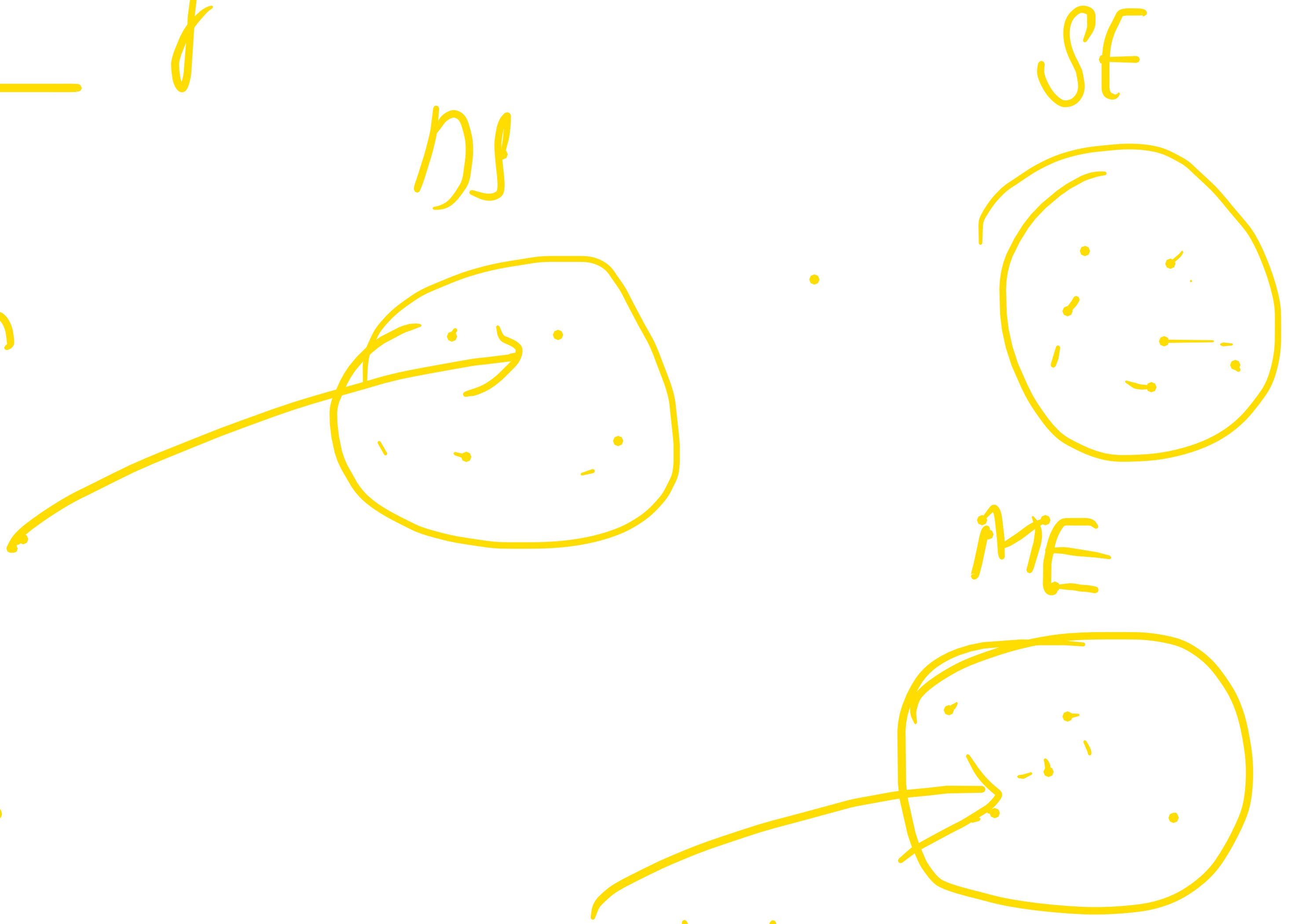
A → Discounts

C → More Discounts +More Ads



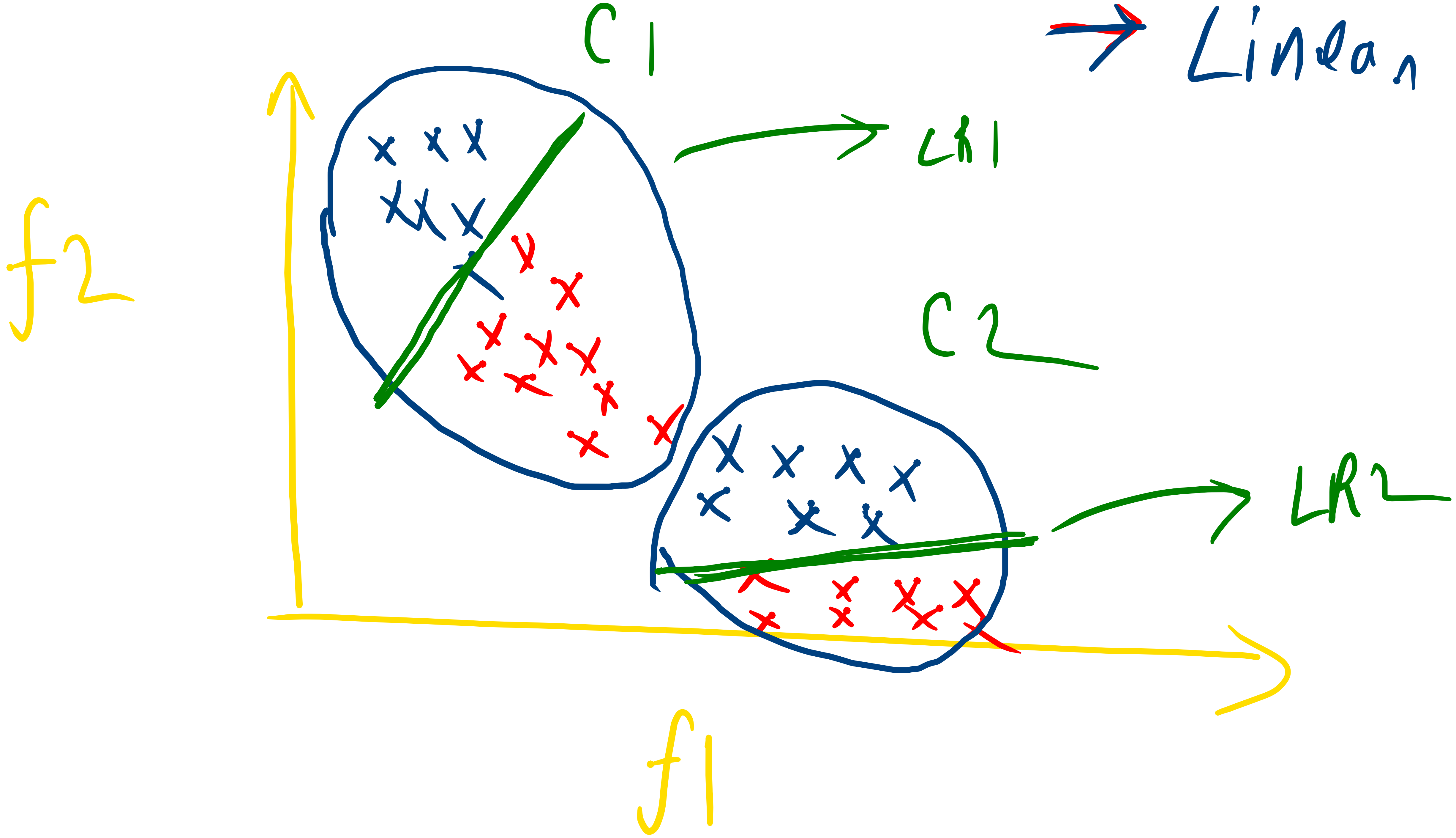
Application of Clustering

→ Search algorithm



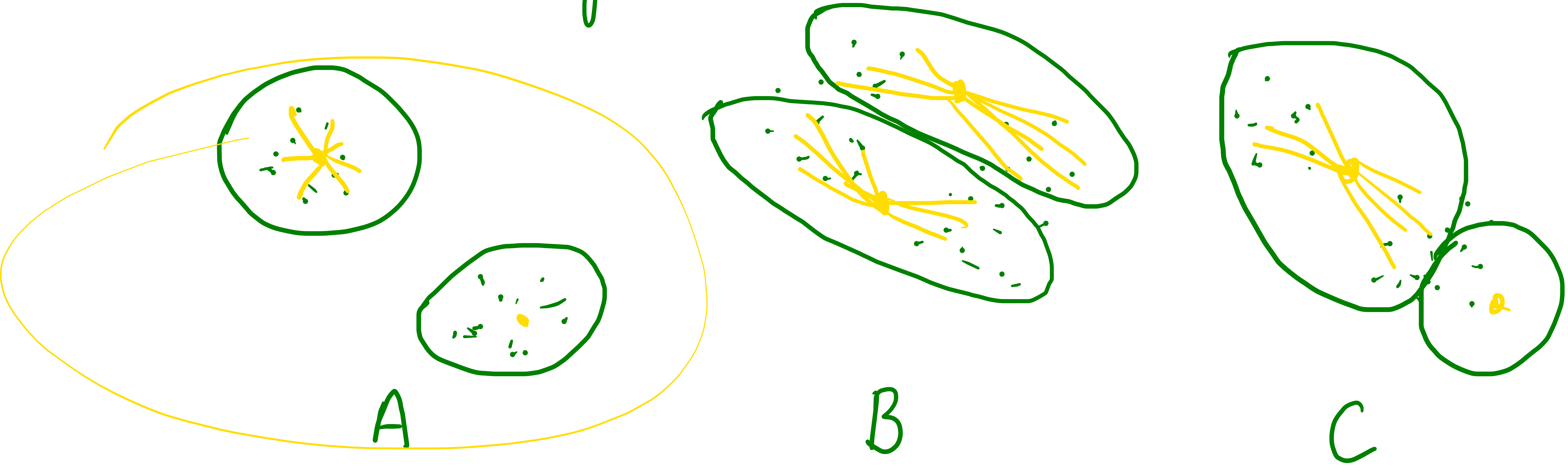
→ Feature creation

ID	n_clicks	n_visit	amount_spent	Country
				1
				2
				2



→ Linear Desicion boundary

Good Clustering



$\underbrace{\rightarrow \text{Intra cluster dist}}$ minimize

$\underbrace{\rightarrow \text{Inter cluster dist}}$ maximize

WCSS (Within Cluster Sum of Square)

$$\text{WCSS} = \sum_{P_i \text{ in } C_1} \text{dist}(P_i, C_1)^2 + \sum_{P_i \text{ in } C_2} \text{dist}(P_i, C_2)^2 + \dots + \sum_{P_i \text{ in } C_j} \text{dist}(P_i, C_j)^2$$

$$\text{WCSS} = \boxed{\sum_{k=1}^{K} \sum_{i=1}^{n} 1(c_i = k) \, \| x_i - \mu_k \|^2} = L$$

$$= \boxed{\sum_{k=1}^{K} \sum_{i \in C_k} \| x_i - \mu_k \|^2}$$

$C_1 \quad C_2$

Gradient Descent

Update all parameters simultaneously

$$W \leftarrow W - \alpha \nabla_W L$$

↳

w_1

w_2

$\vdots$

w_n

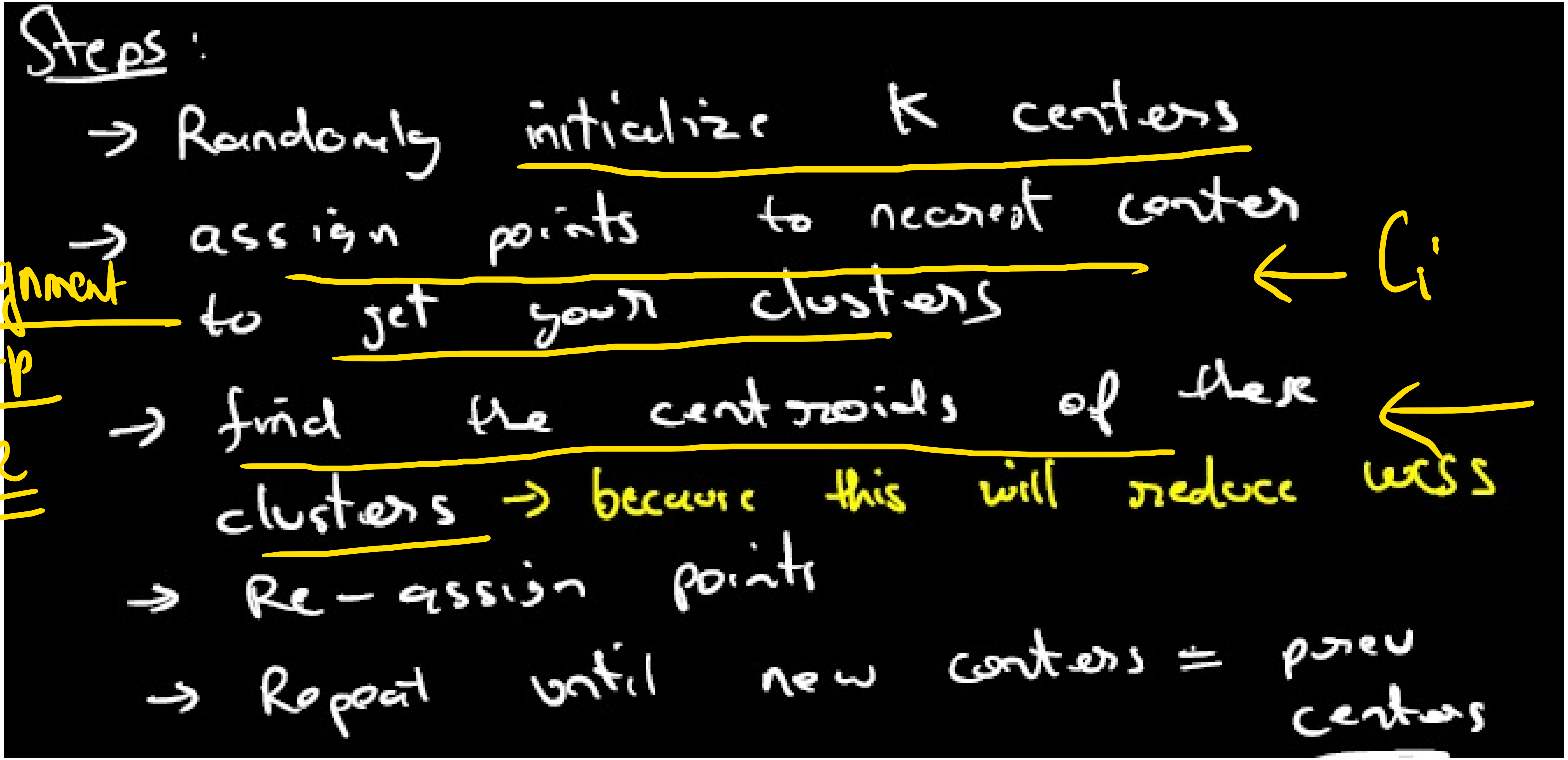
Coordinate Descent

→ Update only a subset of parameters

↳ Fixing μ & we find best c exactly

↳ Fixing c & find μ

K-means Clustering (Lloyd's Algorithm)



Steps:

→ Randomly initialize K centers

Assignment Step

→ assign points to nearest center to get your clusters ← C_i

Update

→ find the centroids of these clusters → because this will reduce WCSS ← μ_K

→ Re-assign points

→ Repeat until new centers ≐ prev centers

Red cluster

$(1, 2)$

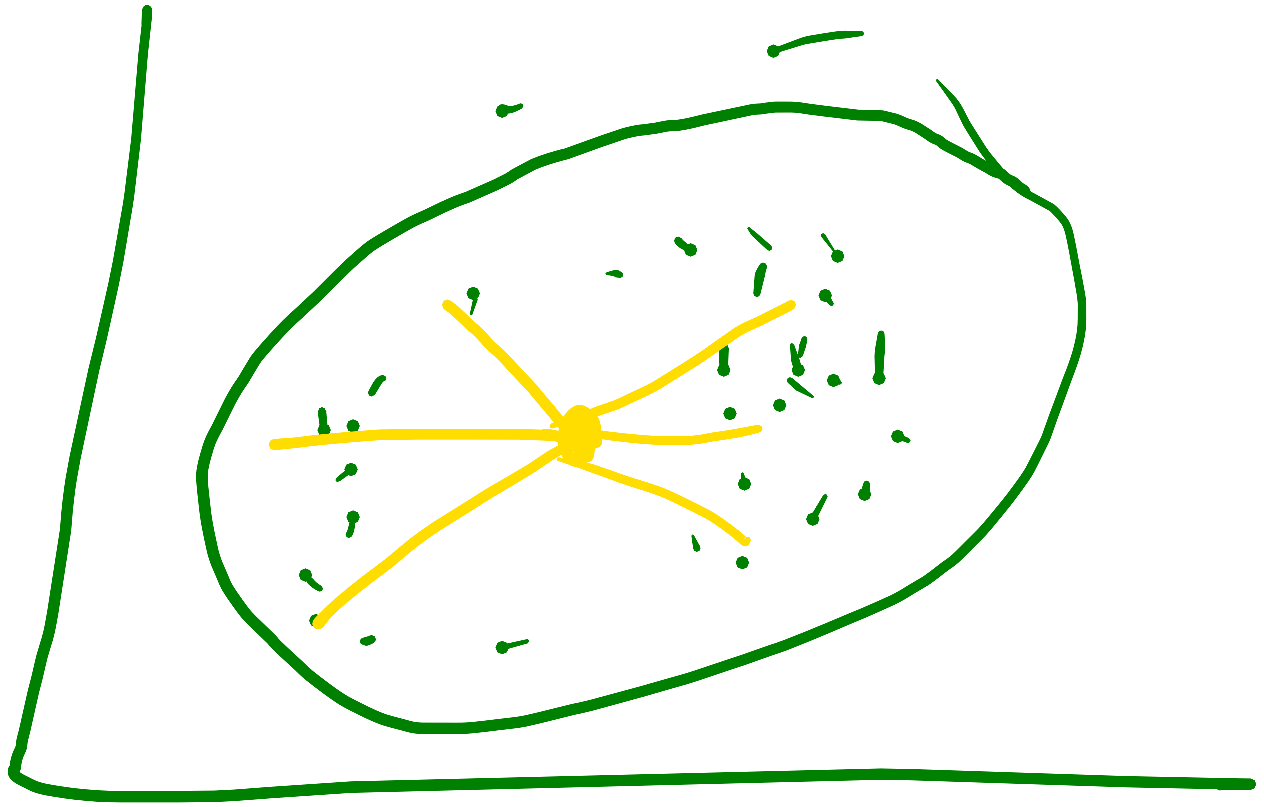
$(2, 3)$

$(3, 4)$

$\left(\cancel{6}/3 ,\ \cancel{9}/3 \right) = (2, 3)$ ← Centroid

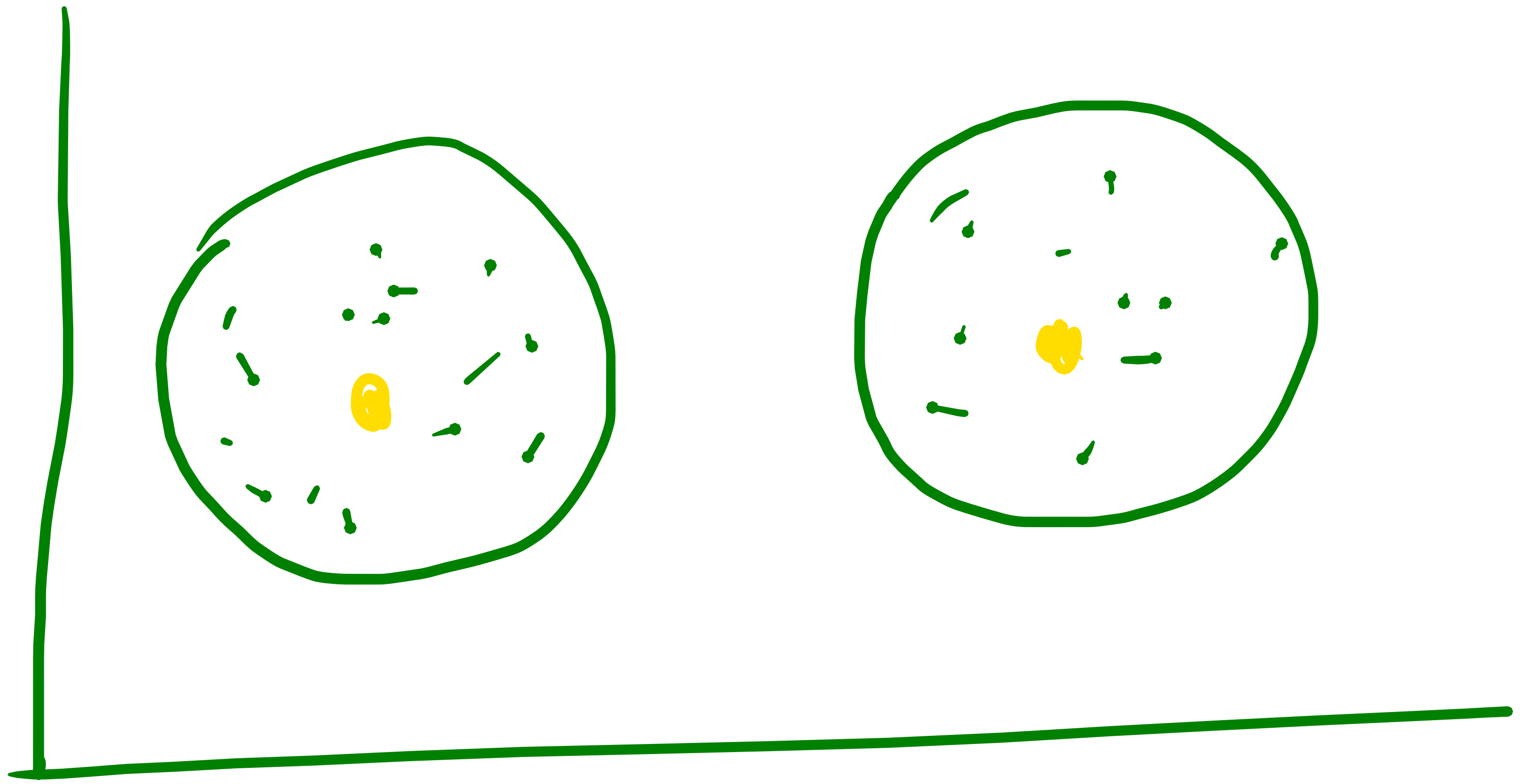
✓ $$\boxed{\mu_K = \frac{1}{n_K} \sum_{\underline{x_i \in C_K}} x_i}$$

$\mu_{new} - \mu_{old} < 10^{-4}$



Cluster = 1

↑



Cluster = 2

↓ WCSS ↓

→ Break until 22:33

Busines

→ No of Ads

→ Amount of Discount