center of cluster = the mean of samples,
not nesesery a sample

grouping into # of groups → hyperparameter

* clustering is nondetermenistic (not same solution always)

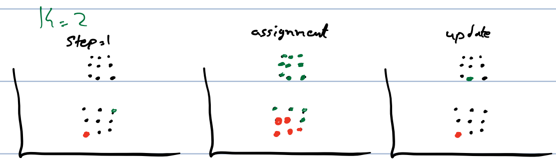
loss fun → * measure ⇒ Σ of squared distances from center (smaller is better)

3 technices of clustring:-

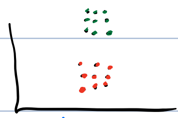
depends on the initial assignment

① K-means (NP-hard)

→ # of clusters (hyper-parameter)



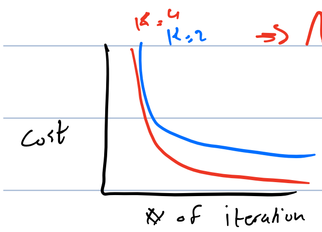
① select K centers (nearst by Euclidea)



② loop { assignment to nearst K
update the center

NOT always optimal

③ converge ⇒ no chang in step 2 always converge



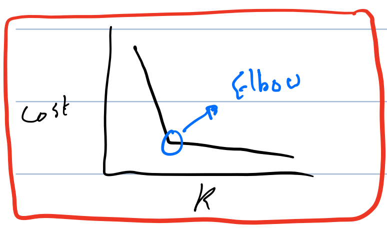
⇒ NOT every Solution

no error ⇒ K = # of Samples

How to find K ?

- from your problem (# of room ...)

- Elbow technick



→ more K ⇒ less cost

K-means ++ (Smart Initialization)

to ensure you select good centers

Select first center randomly, then select the farest random

non-deterministic:-

1- multiple runs

2 - Smart run K-means++

types of K-means

① LBG algorithm ⇒ Start with 2 then split them untill we got K

② MacQueen algorithm

Speed: update as you assign faster

good: close to normal K-means

problems:-

- k-means is spherical because we use L_2 distance
- if you have overlapping
- all clusters are equals

expectation maximization

it has fixed solution (more roubost)

② EM for mixtures of Gaussian → soft clustring

- not sphericals
- not identified by center, identified by Gaussian PDF (mean and σ^2)

normal distribution

we will have 2 means in 2D plane

will be covariance

Calculate the probability of a point that belongs to N_{blue}

blue cluster

$$P(y=b \mid x, N(\mu_b, \sigma_b)) = \frac{P(x \mid y)\; P(y)}{P(x)}$$

π_j

w_{ij} — Probability of sample x belongs to blue cluster

is 1D

$$\Rightarrow \frac{1}{\sqrt{2\pi}\,\sigma_b} e^{-\frac{1}{2}\frac{(\mu_b - x)^2}{\sigma_b^2}}$$

the maximum Probability

more than 1D

$$\frac{1}{\sqrt{(2\pi)|\Sigma|}^{1/2}} e^{\frac{1}{2}\frac{(\vec{\mu}_i-\vec{x})^T(\vec{\mu}_i-\vec{x})}{\Sigma^2}}$$

$$\Sigma = \begin{bmatrix} \sigma_1 & cov_{12} \\ cov_{21} & \sigma_2 \end{bmatrix}$$

Perfectly -ve corr — Perfectly +ve corr

$$\text{Correlation} = \frac{cov}{\sigma_1 \sigma_2} = [-1, 1]$$

if corr between two features is 1 you can remove one of them

$$Cov(x, y) = \mu(xy) - \mu(x) * \mu(y)$$

How to cluster:-

① randomly select 2 samples (= # of distrebutions)

② μ = sampls , σ = random

③ $P(y) = \frac{1}{2}$ (π_j)

Slide UO ← ④ loop

- expectation (assign) → all samples
- maximization [update the means (wieghts* mean)]

- both may stuck in local minima
- EM is not better than K-means in all problems

K-means

- initialize the cluster

$c_1 = \mu_1 = x_i$

$c_2 = \mu_2 = x_i$

$\vdots$

μ_K

parameters = # of features

- Repeat {
  - assign
  - update : $c_i = \frac{x_1 + x_2 \cdots}{n_i}$

}

- converge = no new assign

EM

- initialize the clusters (Gaussian)

μ_i, $\Sigma_i^{d \times d}$ (σ^2), $\pi_i^{\frac{1}{K}}$ $(P(y=k))$

- Repeat {
  - expectation: $\frac{P(x|y_i) P(y_i)}{\sum_{j=1}^{K} P(x|y_j) P(y_j)} = P(y|x)$ = w_{ij} (sample, cluster)
  - maximization: $\mu_j = \frac{\sum_{i=1}^{m} x_i w_{ij}}{\sum_{i}^{m} w_{ij}}$ (samples)

    $\pi_j = \frac{\sum w_{ij}}{m}$

}

	$K=1$	$K=2$
S_1	0.4	0.6
S_2	0.9	0.1
S_3	0.8	0.2

$\mu_1 = \frac{7 \times 0.4 + 7 \times 0.9 + 0.8}{2.1}$

	x_1	x_2	$x_1 x_2$
S_1	6	4	24
S_2	3	3	9
S_3	4	3	12

$Cov(x_1, x_2) = E(x_1 x_2) - E(x_1) E(x_2)$

$= 15 - 4.3 \cdot 3.3 \approx 1.0$

$\mu_1 = [\mu_{11}, \mu_{12}]$

$\pi_{K=1} = \frac{w_{i1}}{n_1}$

$\Sigma_{K=1} = \begin{bmatrix} \sigma_{11}^2 & Cov(12) \\ Cov(12) & \sigma_{21}^2 \end{bmatrix}$

- converge when the change is small

③ Agglomerative hierarchical clustring

- no need for input K

- clustering at multiple levels

- bottom up (looking for smallest) Dendrogram (threshold is hirozental line) should use the biggest gap

※ measuring the distance between clusters:-

- Single linkage algo Smallest linkage (chains)

- complete linkage algo Largest linkage (homognized)

- Ward algo average

- **Measuring the distance between to clusters:**

Now the question is, how do we measure the distance between two clusters? We will explain to methods:

1-Single linkage: in this method, we will take the distance between the closest two points in two clusters and we will consider it as the distance between the two cluster. For example, suppose we have two cluster, C1 has two sample (X1 , X2) and C2 has two samples (X3 , X4). Assume the distance between them is as follow:

$Distance(X1, X3) = 2$

$Distance(X1, X4) = 1.5$

$Distance(X2, X3) = 7$

$Distance(X2, X4) = 0.5$

Since the smallest distance 0.5 then we will consider the distance between C1 and C2 is 0.5.

2- Complete linkage: in this method, we will take the distance between the furthest two points in two clusters and we will consider it as the distance between the two cluster. If we take our previous example and using the complete linkage method, then the distance between C1 and C2 will be 7.

-We can use Euclidean distance two measure the distance between two points.

-After we found all the distance between clusters, we will merge the closest two clusters as we did in our previous example.