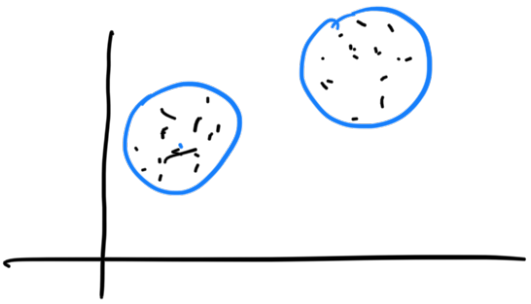
CO2774

Machine learning

Oct 22, 2021

Unsupervised learning :- { Clustering
Density Estimation
PLA

Clustering :- K-Means Algorithm



$\{x^{(i)}\}_{i=1}^{m}$

⇒ (A) $\mu_1 - \mu_K$ (cluster means)

(B) $\{c^{(i)}\}_{i=1}^{m}$ cluster assignments

$\underset{\{\mu_k\}_{k=1}^{K}, \{c^{(i)}\}_{i=1}^{m}}{\text{argmin}} \sum_{i=1}^{m} \|x^{(i)} - \mu_{c^{(i)}}\|^2$ — (1)

↳ (A) If we knew μ_k's, can compute $c^{(i)}$'s

↳ (B) If we knew $c^{(i)}$'s, can compute μ_k's

do {

$c^{(i)} = \underset{k}{\text{argmin}} \|x^{(i)} - \mu_k\|^2$ — (A)

$\mu_k = \underset{\mu_k}{\text{argmin}} \sum_{i=1}^{m} 1\{c^{(i)} = k\} \|x^{(i)} - \mu_k\|^2$ — (B)

} (while ! converged)

⇒ K-Means Algorithm :-

(1) start with an initial estimate of cluster centroids

Given :- $\{x^{(i)}\}_{i=1}^{m}$ $x^{(i)} \in \mathbb{R}^n$

# of clusters :- K

$\{\mu_k\}_{k=1}^{K} \leftarrow$ init();

② Repeat{

$\forall i: \; c^{(i)} \leftarrow \underset{k}{\text{argmin}} \; \|x^{(i)} - \mu_k\|^2$ — (A)

$\forall k: \; \mu_k \leftarrow \underset{\mu_k}{\text{argmin}} \sum_{i=1}^{m} 1\{c^{(i)}=k\} \|x^{(i)} - \mu_k\|^2$ — (B)

} (while ! converged)

⇓ H.W.

$$\mu_k = \frac{\sum_{i=1}^{m} 1\{c^{(i)}=k\} x^{(i)}}{\sum_{i=1}^{m} 1\{c^{(i)}=k\}}$$ ✓

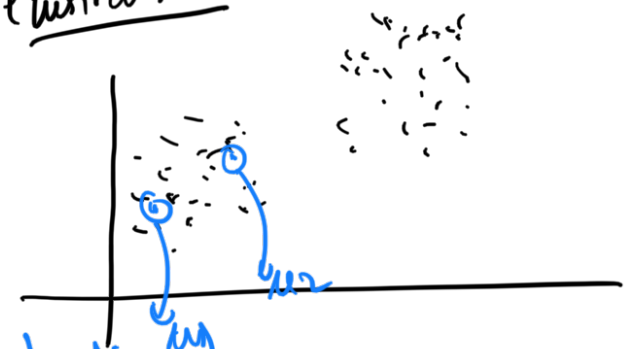
How to detect converged?

↳ If $c^{(i)}$'s are not changing
or equivalently μ_k's don't change.

"local minima"

illustration:

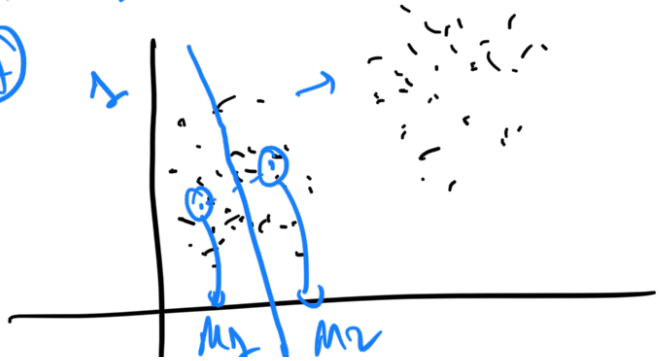
$K=2$ $\{x^{(i)}\}_{i=1}^{m}$
$x^{(i)} \in \mathbb{R}^2$



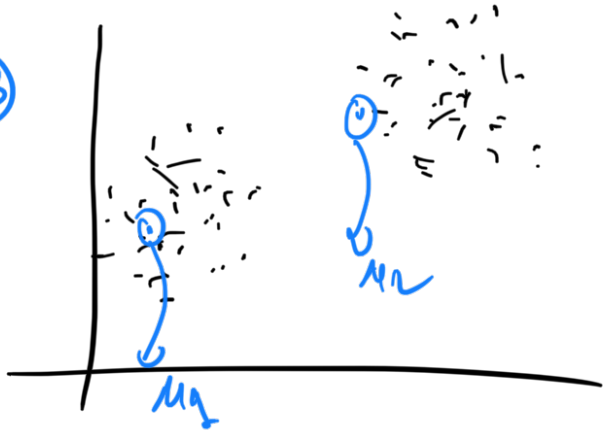
initialization

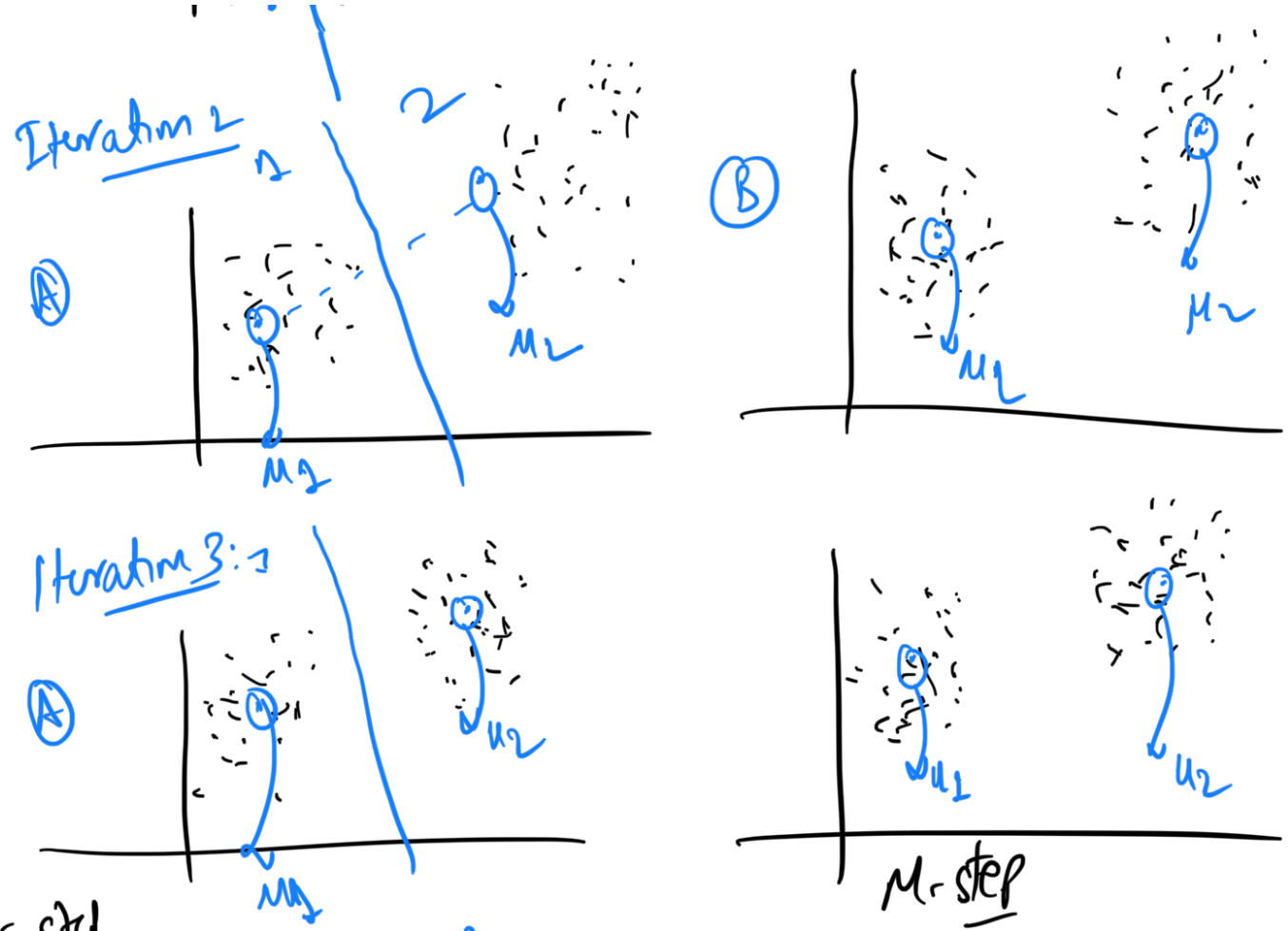
Iteration 1:

(A)



(B)





Convergence.

Convergence: "Various" settings (local optima)

"No guarantee" as such

Time complexity:- $O(T \cdot (n \cdot k \cdot n + n \cdot m))$ → $\sum_{k=1}^{K} n \cdot m_k$

$\equiv O(nmkT)$

E-step

M-step

$x^{(i)} \in \mathbb{R}^n$

K

m:- # of points

Computing $\mu_1 \ldots \mu_k$ ... $n \cdot m_k$

⇓

# of points assigned to cluster k

$m_k = \sum_{i=1}^{m} 1\{c^{(i)} = k\}$

K-Means can be seen as

Optimization of ... 0

objective function:-

$$\underset{\{\mu_k\}_{k=1}^{K},\{c^{(i)}\}_{i=1}^{m}}{\operatorname{argmin}} \sum_{i=1}^{m} \| x^{(i)} - \mu_{c^{(i)}} \|^2$$

using Block coordinate descent.

⇒ Optimize one block (set) of variables at a time while keeping others fixed.

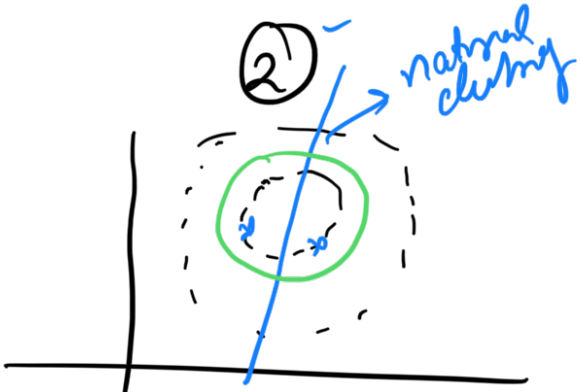
⇒ "Local optima" of the objective function

Finally:- When does K-means algorithm not work?

① ✓

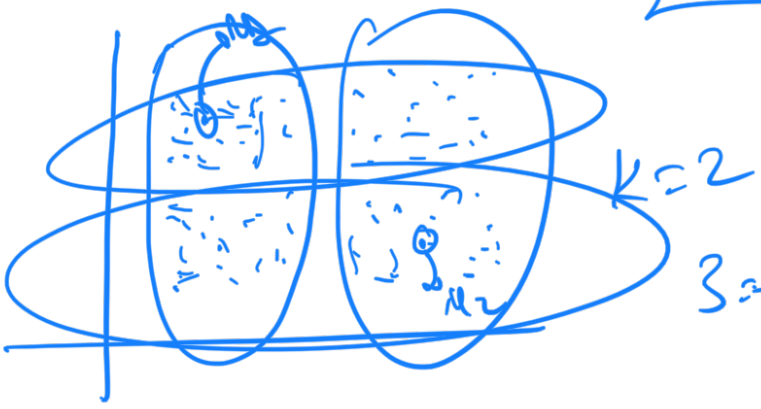
"Assume" clusters of similar in size

② ✓



Assume data points are well separated

GMM (Gaussian Mixture Model)



3:- The final clustering depends on starting point

Next class:- GMM (Gaussian Mixture Models)

GDA:- classification.

$\{x^{(i)}\}_{i=1}^{m}$

$\{x^{(i)}, z^{(i)}\}_{i=1}^{n}$ Hidden

$$p(x^{(i)};\theta) = \sum_{z^{(i)}} p(x^{(i)}|z^{(i)};\theta)\; p(z^{(i)};\theta)$$

↳ Gaussian

Mixture

Audit:-

AND [At least 40%
is at least 30% in theory / exam
and least 30% in practical (assignments)]

OR

equivalent of C/B grade?