ANNOUNCEMENTS

1. HW1 released last saturday, due next Tuesday
2. Numpy tutorial recording posted to Ed
3. [5780] Reading paper out on website, Quiz out tonight/tomorrow

---

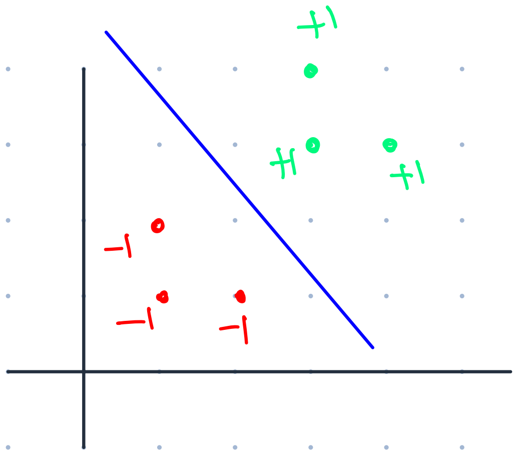
Put your devices away, notes are to the front (see last page for notation!)

we will give you a device break!! :)

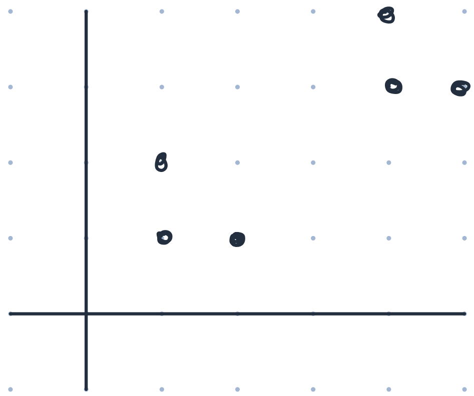
Aside (while you wait): we, ML practitioners often believe in the concept of "god" to make our ML models converge!!

Unsupervised Setting

Today: k-means, mixture of Gaussians



Supervised setting
(INCLUDES class labels)



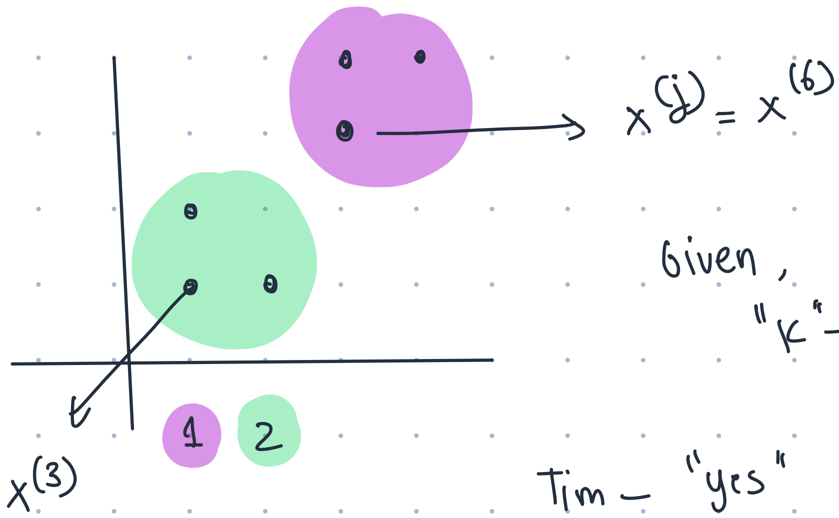
Unsupervised setting
(NO class labels)

Unsupervised setting is "hard"

1. make stronger assumptions
2. accept weaker guarantees!

K-MEANS

j = data index
i = cluster idx
k = # clusters/groups



Given, $x^{(1)} \dots x^{(n)} \in \mathbb{R}^d$
"k" — # clusters

Tim — "yes" because we know from experience
Arker — "no" million data, intractable

<u>Do</u>: find cluster assignment for each point, $x^{(j)}$

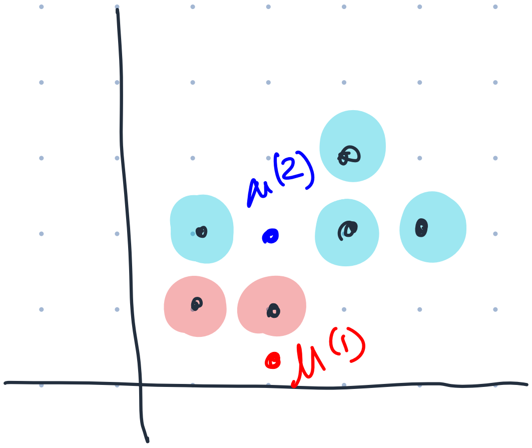
$$c^{(j)} = i$$

E.g., for $x^{(3)}$, cluster-2 ← $c^{(3)} = 2$
$x^{(6)}$, cluster-1 ← $c^{(6)} = 1$

$$1\{c^{(j)} = i\} = \begin{cases} 1, & \text{if } x^{(j)} \text{ belongs to cluster-}i \\ 0, & \text{otherwise} \end{cases}$$

"does $x^{(j)}$ belong to cluster-i"

How do we find these cluster assignments?



1. Randomly assign "k" cluster centers — $\mu^{(i)}$

REPEAT

2. Assign each data point to the "closest" cluster

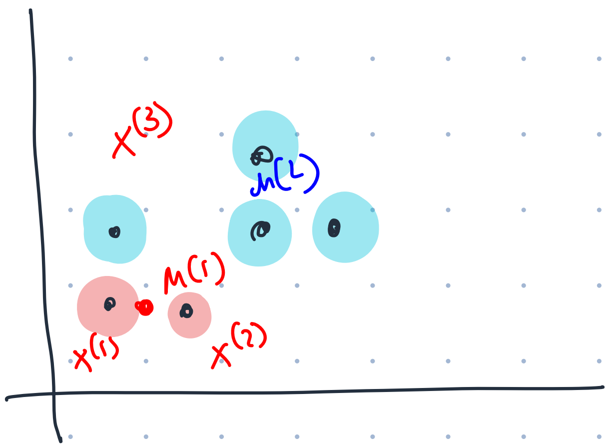
for $i = 1, \ldots, n$

$$c^{(j)} = \underset{i=1\ldots k}{\arg\min} \; \|x^{(k)} - \mu^{(i)}\|^2$$

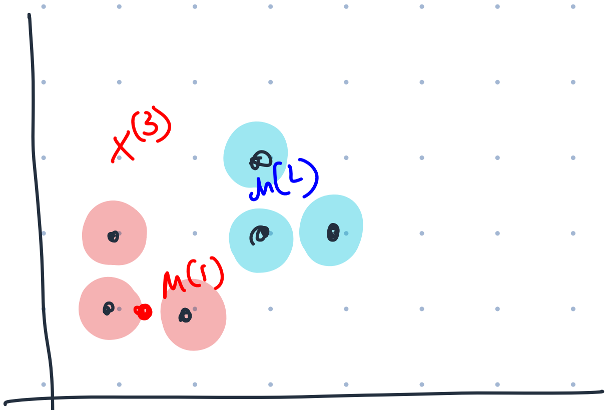
"NP hard" problem — no polynomial time algo

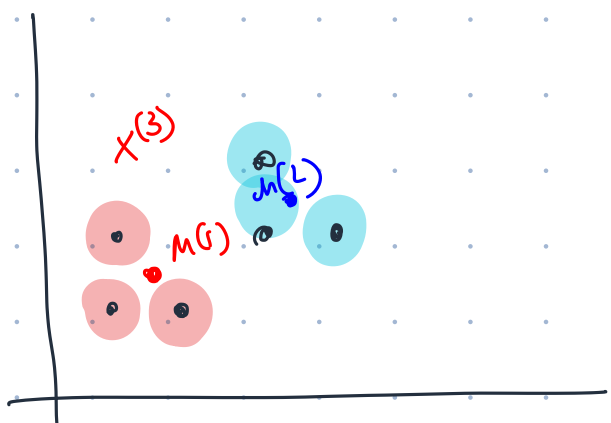
3. Recompute cluster centroids based on the cluster assignments

$$\mu^{(i)} = \frac{\sum_{j=1}^{n} 1\{c^{(j)} = i\} \cdot x^{(j)}}{\sum_{j=1}^{n} 1\{c^{(j)} = i\}}$$



$$\mu^{(1)} = \frac{x^{(1)} \cdot 1 + x^{(2)} \cdot 1 + x^{(3)} \cdot 0 + \ldots}{1 + 1 + 0 + 0 + 0 + 0}$$





"NP-hard" — no polynomial time

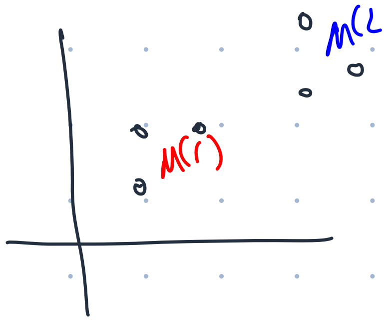
Q. Does it converge — "yes"

Picking cluster assignments — have data points closest to centroid

picking centroid — have centroid closest (average) of data points

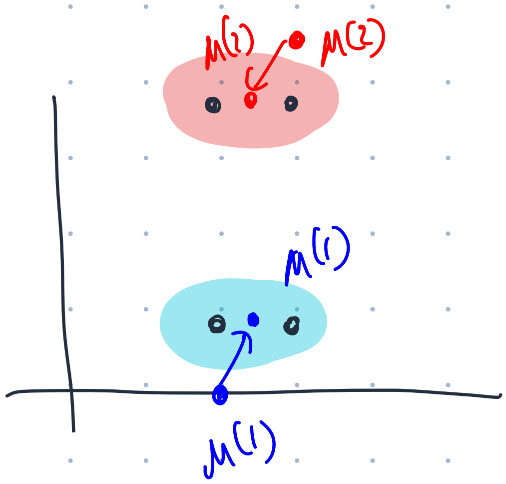
$$J(c, \mu) = \sum_{j=1}^{n} \|x^{(j)} - \mu^{(c^{(j)})}\|^2$$

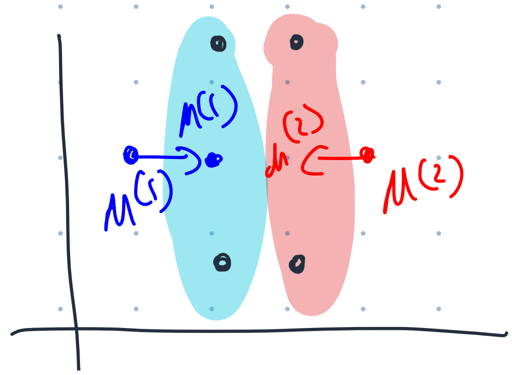
centroid of the cluster datapoint $x^{(j)}$ was assigned to



J always decreases or remains the same!

Q. Will it always converge to the same centers? — No!

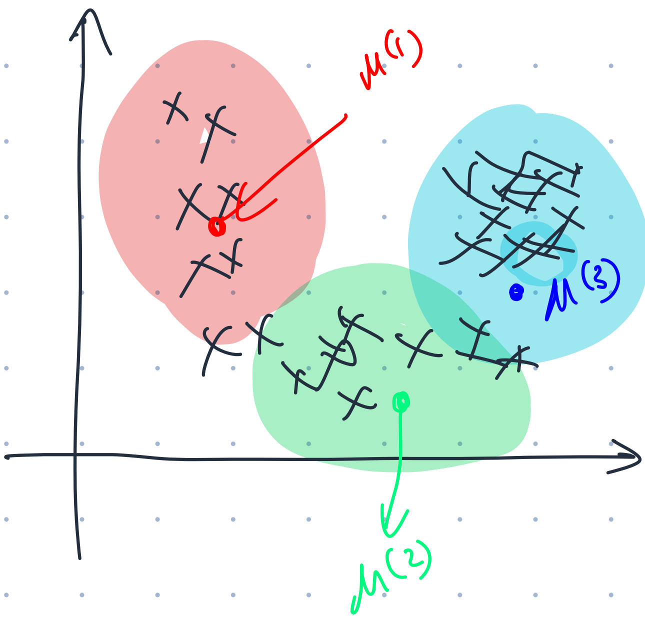




global optimum is not guaranteed!

How to initialize — Pick random data points as centers
Repeat! many times
Pick the optimal clusters based on $\underline{\underline{J}}$

Mixture of Gaussians



3 different light sources (galaxies, stars, qasars)

<u>Do</u>: Assign each data point to a light source

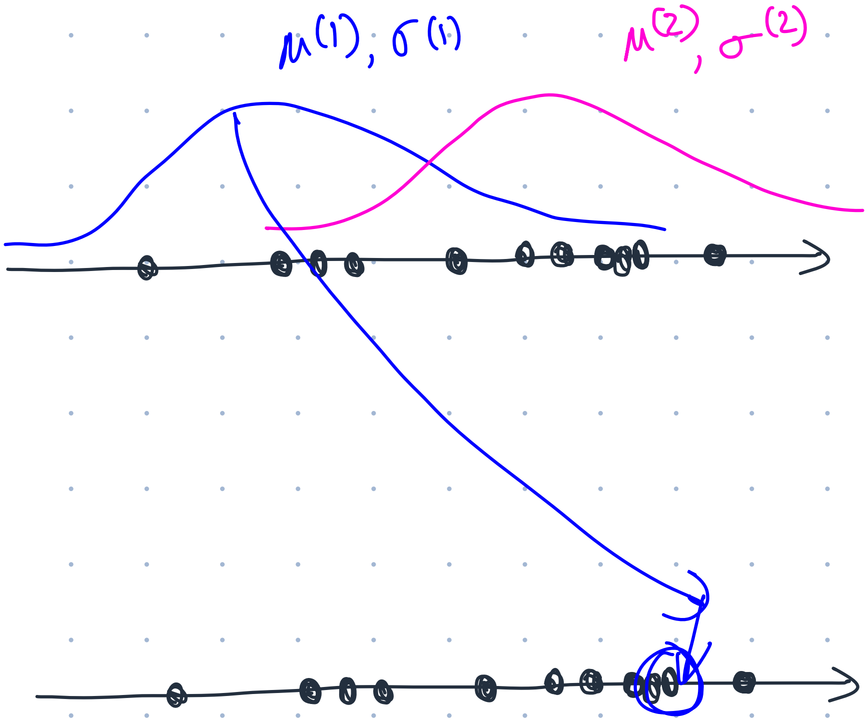
$$p(z^{(j)} = i \mid x^{(j)})$$

" probability that point $x^{(j)}$ came from light source-i, given that we recorded $x^{(j)}$ "

Assumptions:

1. many light sources, we know how many — 'k' known
2. Each light source is well modeled by a Gaussian
3. <u>NOTE</u>: each light source ~~generated same number of photons~~

The mixture of Gaussians



1. pick one of the two gaussians
2. Sample from the chose $\mathcal{N}$

with probability p

$$P(z^{(j)}=1) = p$$

" probability of choosing N_1 is p "

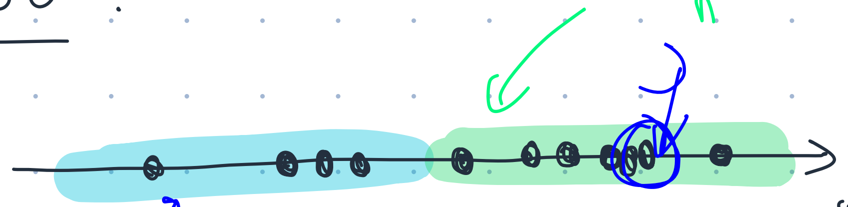
Trying to do: $P(z^{(j)}=i \mid x^{(j)})$

prob of point $x^{(j)}$ ~~belonging to~~ ~~cluster-i~~ was generated from gaussian-i

given we obs $x^{(j)}$

ASIDE :

$\frac{\Sigma}{n} = \mu_{N_2}, \sigma_{N_2}$



$= 4/10$

$\mu_{N_1} = \text{avg} \leftarrow N_1$

$\sigma_{N_1} = \Sigma(x-\mu)^2$

N_2

" probability density"

probability of choosing gaussian - i

BAYES

$$P(z^{(j)} = i \mid x^{(j)}) = \frac{P(x^{(j)} \mid z^{(j)} = i) \cdot P(z^{(j)} = i)}{P(x^{(j)})}$$

$$\boxed{P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}}$$

probability of $x^{(j)}$ given that we picked Gaussian - i

$$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

x

" probability of choosing photon, $x^{(j)}$ "

given the cluster assignments, we can recover gaussians!!

Gaussian mixture model

1. Assume values for $\mu^{(1)}, \mu^{(2)}, \sigma^{(1)}, \sigma^{(2)}, p$

2. Do cluster assignments — $P(z^{(j)}=i \mid x^{(j)})$

prob that $x^{(j)}$ came from N_i given ...

3. Update $\mu^{(1)}, \mu^{(2)}, \sigma^{(1)}, \sigma^{(2)}, p$ based on the cluster assignments (see notes!)

prob of data point — G1 / G2

$$P(x^{(j)}) = P(x^{(j)} \mid z^{(j)}=1) \cdot P(z^{(j)}=1) + P(x^{(j)} \mid z^{(j)}=2)\, P(z^{(j)}=2)$$