LLYOD'S ALGORITHM

QUESTIONS:

1. CONVERGENCE
2. NATURE OF CLUSTERS
3. INITIALIZATION
4. CHOICE OF K.

Note Title 8/29/2022

CONVERGENCE

- Does Llyod's Algorithm Converge? YES

PROOF:

$\boxed{\text{FACT 1:}}$ Let $x_1, x_2, \ldots, x_\ell \in \mathbb{R}^d$

$$v^* = \arg\min_{v \in \mathbb{R}^d} \sum_{i=1}^{\ell} \|x_i - v\|^2$$

[view this objective as a function of v, take derivative, set to 0 and solve]

Answer: $v^* = \frac{1}{\ell} \sum_{i=1}^{n} x_i$

Note Title 8/29/2022

- Say we are at iteration t of Llyod's algorithm.

- CURRENT ASSIGNMENT.

$$Z_1^t, Z_2^t, \ldots, Z_n^t \in \{1, \ldots, k\}$$

$\mu_k^t \leftarrow$ Mean of cluster k in iteration t

- Say we update our assignments to

$$Z_1^{t+1}, Z_2^{t+1}, \ldots Z_n^{t+1} \in \{1, \ldots, k\}$$

Note Title 8/29/2022

$$\rightarrow \underline{\sum_{i=1}^{n} \| x_i - \mu^t_{z_i^{t+1}} \|^2} < \underline{\sum_{i=1}^{n} \| x_i - \mu^t_{\underline{z_i^t}} \|^2}$$

[By Algorithmic choice]

Mean of cluster where x_i wants to go to

$F(z_1^t, \dots, z_n^t)$

Mean of current cluster where x_i is assigned to.

$$\sum_{i=1}^{n} \| x_i - \mu^{t+1}_{z_i^{t+1}} \|^2 \leq \underbrace{\sum_{i=1}^{n} \| x_i - \mu^t_{z_i^{t+1}} \|^2}_{} \leftarrow$$

$F(z_1^{t+1}, \dots, z_n^{t+1})$

Note Title 8/29/2022

$$= \underbrace{\sum_{i=1}^{n} \sum_{k=1}^{K} \|x_i - \mu_k^{t+1}\|^2 \cdot \mathbb{1}(z_i^{t+1} = k)}$$

For every k

$$\sum_{i \in C_k} \|x_i - \mu_k^{t+1}\|^2 \leq \sum_{i \in C_k} \|x_i - v\|^2$$

[FACT 1]

$$\leq \sum_{i \in C_k} \|x_i - \boxed{\mu_k^t}\|^2$$

Note Title 8/29/2022

$\Rightarrow$ The objective function strictly reduces after each re-assignment.

$$F(z_1^{t+1}, \dots z_n^{t+1}) < F(z_1^t, \dots, z_n^t)$$

$\Rightarrow$ There are only "FINITE" number of assignments

$\Rightarrow$ Algorithm must converge!

Note Title 8/29/2022

NATURE OF CLUSTERS

$K=2$

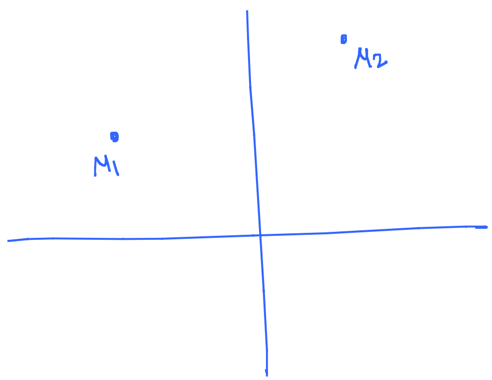
- Llyod's algorithm produces 2 clusters with means μ_1 and μ_2

- What can we say about points assigned to cluster 1 ?

By Algorithm's construction

For points in Cluster 1,

$$\|x - \mu_1\|^2 \le \|x - \mu_2\|^2$$



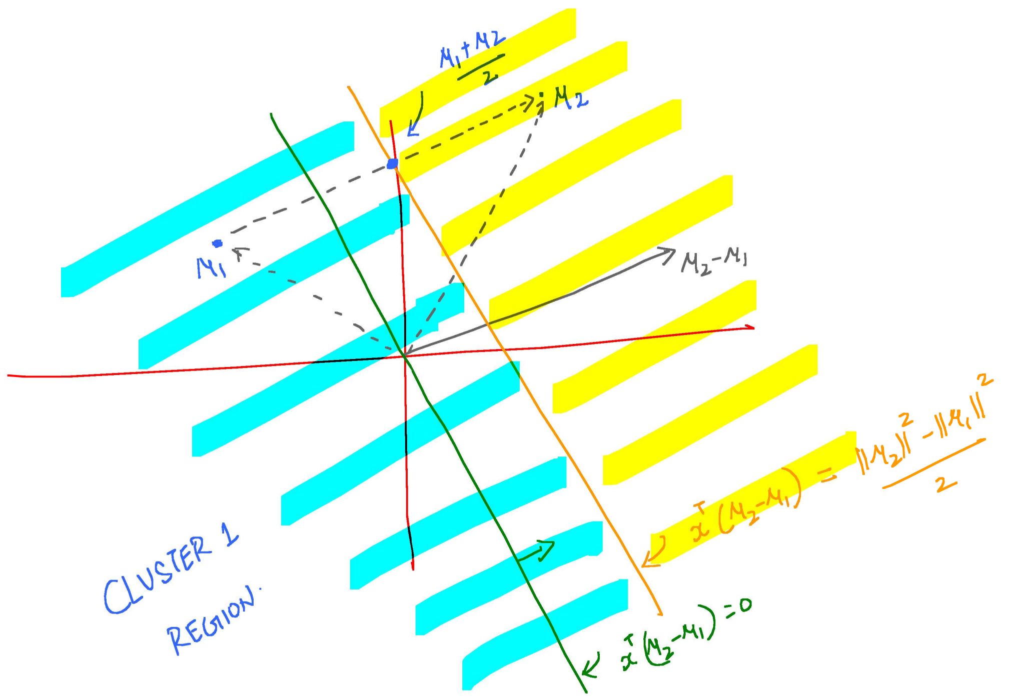
$$\|\cancel{x}\|^2 + \|\mu_1\|^2 - 2x^T\mu_1 \le \|\cancel{x}\|^2 + \|\mu_2\|^2 - 2x^T\mu_2$$

$\forall x$ in cluster

$$\boxed{x^T(\mu_2 - \mu_1) \le \frac{\|\mu_2\|^2 - \|\mu_1\|^2}{2}}$$

$$x^T w \le b$$

Note Title 8/29/2022



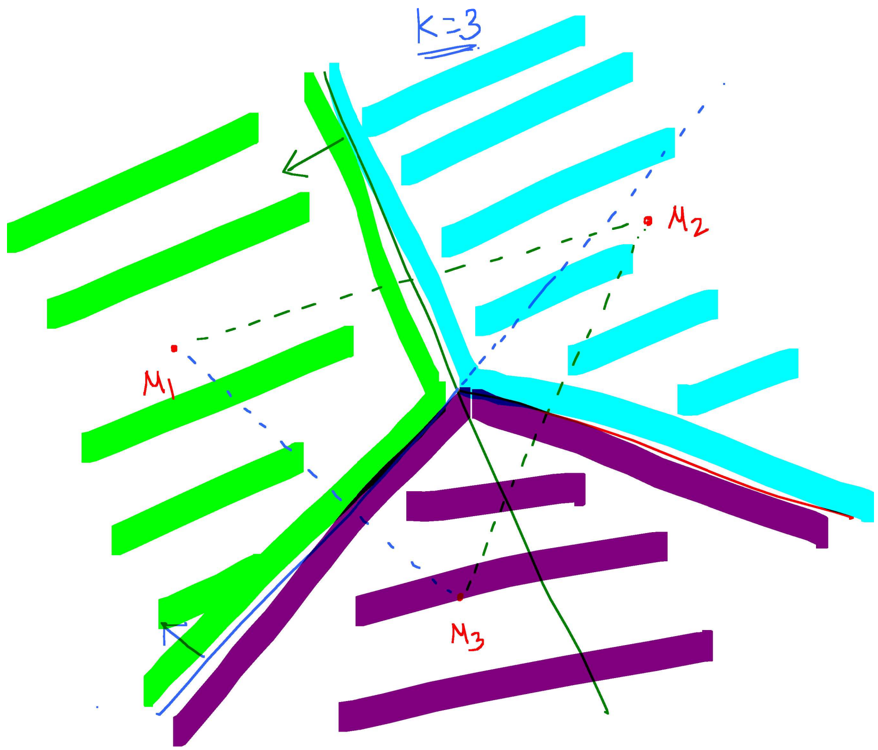
$$\underline{x^T(\mu_2-\mu_1)} \le \underbrace{\frac{\|\mu_2\|^2-\|\mu_1\|^2}{2}}$$

$$x = \frac{\mu_1+\mu_2}{2}$$

$$\left(\frac{\mu_1+\mu_2}{2}\right)^T(\mu_2-\mu_1)$$

$$= \frac{\|\mu_2\|^2 - \|\mu_1\|^2}{2}$$

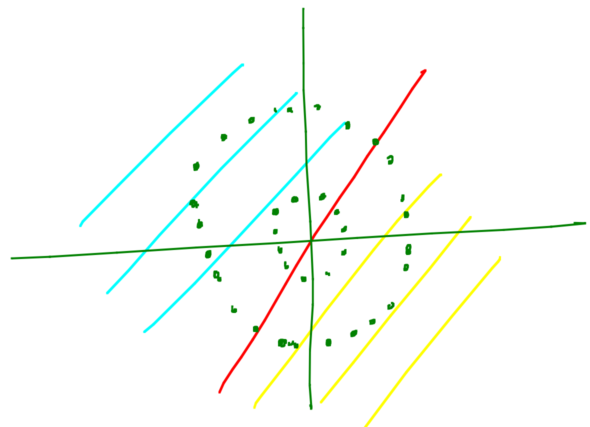
Note Title 8/29/2022



Cluster regions are intersection of Half Spaces

VORONOI REGIONS

Note Title 8/29/2022



How to fix?

- KERNELIZE K-MEANS
- SPECTRAL CLUSTERING.

Note Title 8/29/2022

INITIALIZATION

POSSIBILITIES

- Pick K- means uniformly at random from the dataset

K- MEANS ++

iteration

→ Choose first mean μ_1^0 uniformly at random from $\{x_1, \ldots, x_n\}$

CHOICE OF K

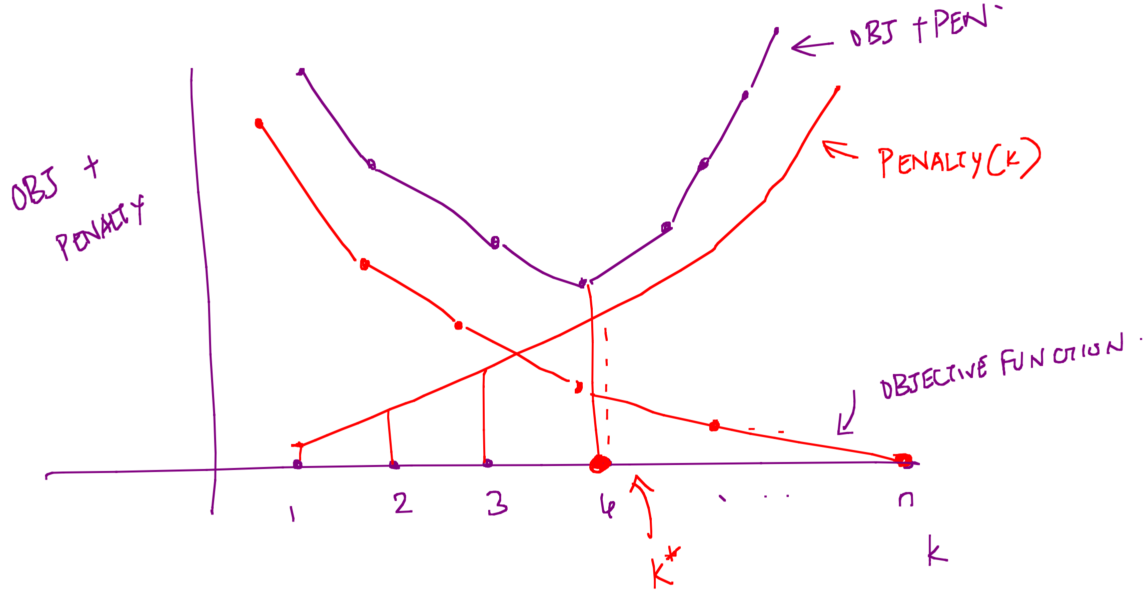
→ $F(z_1, \dots, z_n) = \underbrace{\sum_{i=1}^{n} \|x_i - \mu_{z_i}\|^2}$ $\quad k = n$

→ want K to be as small as possible.

→ Penalize large values of K.

Find k that has the

Smallest objective function value + Penalty(k)



Some common Criterion

A.I.C - Akaike Information Criterion

$$\left[\underline{2K} - \underbrace{2\log\left(L(\theta^*)\right)}\right]$$

B.I.C - Bayesian Information Criterion

$$\left[\underline{K\log(n)} - \underbrace{2\log\left(L(\theta^*)\right)}\right]$$

- CONVERGENCE - YES
- NATURE OF CLUSTERS - VORNOI REGIONS
- INITIALIZATION - K-MEANS++
- CHOICE OF K - OBJ + PENALTY(K).

Note Title 8/29/2022

$\rightarrow$ for $\ell = 2, \dots, k$

Choose μ_ℓ^0 probabilistically proportional to score

$$\forall x \quad S(x) = \min_{j=1,\dots,\ell-1} \| x - \mu_j^0 \|^2$$

$x_1, x_2, \boxed{x_3}$

$10 \quad 20 \quad \boxed{30}$

$1/6 \quad 1/3 \quad \boxed{1/2}$

GUARANTEE

$$\mathbb{E}\left[\sum_{i=1}^{n} \| x_i - \mu_{z_i} \|^2 \right] \le O(\log k) \left[\min_{z_1,\dots,z_n} \sum_{i=1}^{n} \| x_i - \mu_{z_i} \|^2 \right]$$

over randomness of algorithm