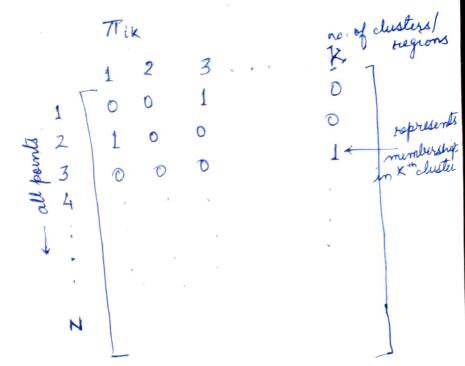
HW-2

PROBLEM - 1 K-Means Objective: min $\sum_{i=1}^{N} \frac{\pi_{ik}}{|X_i|^2} = \frac{\pi_{ik}}{|X_i|^2}$

⇒ variance within each



 (\mathbf{A})

In order to prove that the E-step update on memberships It achieves the minimum objective given the current centroids (u), we must show that reassigning memberships of each point to the newest centroid minimizes the K-means objective function. For each dalapoint, the objective function yields

a single term 11xi- ull2

Therefore, minimizing the objective function means reducing the value of ||xi-u|| for each dalapoint There is only I term because each datapoint can only belong to one cluster. Therefore, when the centroids, U, are fixed, we must reasign memberships (T) of each datapoint to the closest rentroid. $K = \arg\min_{k} ||x_i - \mu_k||^2.$ $T_{ik} = \begin{cases} 1, & i \leq k = arg min ||x_i - \mu_k||^2 \\ 0, & \text{for all other } k \end{cases}$ John this process is repeated for all dalapoints; the total objective function becomes: obj = $\sum_{i=1}^{\infty} \min ||x_i - \mu_K||^2$ Since each xi is assigned to the cluster with

Since each xi is assigned to the cluster with the closest centroid, the overall objective is minimized for the given centroids μ_{K} .

(B)

We must show that updating each centroid to the mean of the datapoints in its chuster minimizes the objective function with respect to the centroids.

Focussing on a single cluster K:

obj $K = \sum_{i=1}^{n} ||x_i - \mu_K||^2$

To minimize the objective function, we must find a optimal Ux. This is because the function is exponential & not linear. Larger distances are penalized more heavily compared to shorter distances.

obj $k = \sum_{x_i \in k} [(x_i - \mu_k)^T (x_i - \mu_k)]$

 $= \sum \left[\|x_i\|^2 - 2x_i^T \mu_k + \|\mu_k\|^2 \right]$

since $||\chi_i||^2$ is independent of \mathcal{U}_k , it is a constant. Thus we can focus on minimizing:

obj $x = -2 \sum_{x_i \in k} x_i^T \mu_x + |\mathcal{E}_k| ||\mathcal{U}_k||^2$ no. of points in

2 Mr

set derivative to 0 to find the local

$$-2\sum_{x_i \in k} x_i + 2|k| \mu_k = 0$$

$$\frac{\mu_{k}}{|k|} = \sum_{\substack{\alpha \in K \\ |k|}} \alpha_{i} \in K$$

$$\frac{1}{2i \in K} = \sum_{i=1}^{N} \frac{x_i}{x_i} (K) = \sum_{i=1}^{N} \frac{x_i}{x_i} K$$

$$\frac{1}{2i \in K} = \sum_{i=1}^{N} \frac{x_i}{x_i} K$$

- The E-slip and M-slep allernatively reduce the K Means objective function
 - 1. The E-slep assigns each point to the el nearest senter, which either reduces-the obj func or keeps it the same.
 - 2. The M-step, updates the centroids, which minimizes the (distance) between the points and centroids for each cluster.

Neither step increases & obj.

and since both steps diminish, obj, they must converge to a finite value.

There are A N datapoints & each of them can belong to I cluster only.

Max number of unique cluster combinations is K^N .

Each iteration over E & M step moves to a new configuration that reduces J, so the configuration must eventually reach a local minima

Why K-Means does not necessarily converge to

the global minimum is 0 when - the

mumber of centroids is equal to the

number of datapoints.

Therefore for any all other value of K,

moreover, initial selection of centroids may lead to a poor income.

