Machine Learning: HW1

Brent Garey

February 3rd, 2025

Contents

1 KMeans Theory:

Given Kmeans Objective discussed in class with Euclidian distance:

$min \sum_i \sum_k \pi_{ik} \cdot ||X_i - \mu_k||^2$

A) prove that E step update on membership (π) achieves the minimum objective given the current centroids(μ)

answer: During the E step, we are given current centroids $(\mu) = \{\mu_1, \mu_2 ... \mu_k\}$

Goal: Calculate the memberships $(\pi_{ik}) = \{\pi_1, \pi_2 ... \pi_{ik}\}$ such that the result is the best possible memberships closest to center (μ_k).

The objective of minimizing the objective gets affected by changing the memberships, (π_{ik}).
The vector π_i can only have 1 membership at a time, otherwise having multiple memberships results in doubling the cost (at the least).
Since (π_{ik}) can only take on the values {0,1}, set $\pi_{ik} = 1$ if k = $argmin_{t=1:k}$ $\|x_i - \mu_t\|$.
In the case that it is NOT the argmin, that is, if μ_t is NOT the closest center, then $\pi_{ik} = 0$
Looking at the equation, there are 2 cases when changing π_{ik}:

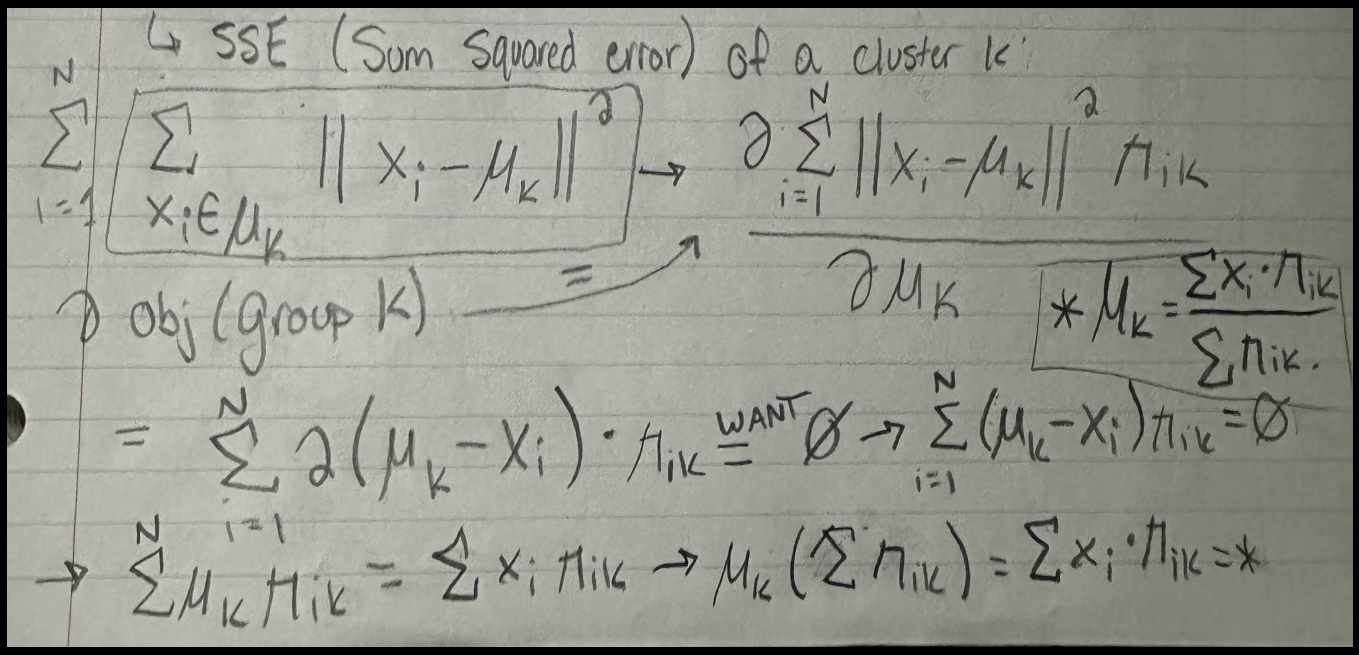
1. Changing a π_{ik} from 0 to 1 when μ_t is NOT the argmin.
2. Changing a π_{ik} from 1 to 0.

Minimizing cost means $\|X_i - \mu_k\|^2 - \|X_i - \mu_t\|^2$ must be > 0 for k to be changed so that μ_t is now the nearest center
π_{ik} essentially acts as a filter, where the $\cdot$ operator provides 0 to the overall cost when $\pi_{ik} = 0$, and (eventually) has to be the minimum when $\pi_{ik} = 1$, otherwise there is a better, minimum cost.

B) prove that M step update on centroids (μ) achieves the minimum objective given the current memberships(π)

answer: During the M step, we are given current memberships $(\pi) = \{\pi_1, \pi_2 ... \pi_{ik}\}$
Focusing on the euclidean distance of $\|X_i - \mu_k\|^2$:
μ_k = avg mean $\{x_i | x_i \epsilon \mu_k\}$ when $\pi_{ik} = 1$.
Minimizing in this step means that we want avg mean $x_i \rightarrow \mu_k$.
This results in 2 cases:
1. mean = 0 $(x_i - \mu_k = 0)$, then x_i must be directly on top of μ_k.
2. higher mean $\rightarrow$ data points further away from μ_k results in a higher variance, up to a point where x_i is closer to another μ_k, which is a contradiction since this would change the membership.

C) Explain why KMeans has to stop (converge), but not necessarily to the global minimum objective value.



answer: looking at the result of the derivation (with respect to μ_k), $= \frac{\sum x_i \pi_{ik}}{\sum \pi_{ik}}$, it is dependent on the memberships in response to x_i. The global minimum objective changes according to μ_k, or the number of centers/clusters. This a local minima, where alternating EM steps will only find the local minima. To achieve the global minima, you CAN increase the number of k clusters, but it becomes arbitrary when you can just set the number of k to the number of data points, leading to overtraining.