Ashutosh Kumar Singh 19CS30008

Q8. Input :- $x_1, x_2, \ldots, x_N \in R^n$,

Initial list of K cluster representatives

$z_1, z_2, \ldots, z_K$.

Output :- Cluster assignment $c_1, c_2, \ldots, c_N$.

(a) Cluster assignment based on cluster representatives.

for each x_i $(i = 1, 2, \ldots, N)$, we need to find

$\|x_i - z_1\|_2, \|x_i - z_2\|_2, \ldots\ldots, \|x_i - z_K\|_K$

For computing $\|x_i - z_1\|_2$, no. of subtractions $= n$, norm $\simeq 2n$, and there are K such computations. This amounts to $(3nK)$.

Then, we also need to find the minimum among all K, that involves K computations.

So, cluster assignment for one x_i takes $(3n+1)K$ operations, and we have N such elements. So, total computational complexity $= \boxed{(3n+1)KN}$.

Ignoring the +1 and the constant factor of 3, we can say this is $\boxed{O(nKN)}$, considering the big-O notation.

(b) update cluster representatives.

For all $j = 1, 2, \ldots, K$

do $z_j = \frac{1}{|G_j|} \sum_{i \in G_j} x_i$

The number of additions will be around $N-K$, but as $K \ll N$, we can say no. of additions $\approx N$

Each addition requires n operations.

So, computations for addition $= nN$ (dividing an n-length vector takes n operations)

After that, number of divisions $= K * n = Kn$

So, total computational complexity for this step $= \boxed{nN + Kn} = \boxed{O(n(N+K))}$

(c) Combining step 1 and 2, we can say that complexity $= nKN + nN + Kn$

As $K \ll N$, and considering big-O notation, we can write,

complexity $\approx O(nKN)$.

So, in 10 iterations, no. of computations $\approx \boxed{10nKN}$.

However, if we want a more exact and rigorous bound, we can calculate that using:-

$$\Big[\underbrace{(3n+1)KN}_{\text{step 1}} + \overbrace{\underbrace{(N-K)n}_{\substack{\text{no. of additions}\\ \text{(exact)}}} + \underbrace{Kn}_{\substack{\text{no. of}\\ \text{divisions}}}}^{\text{step 2}}\Big] \times 10$$

$$= (3nKN + KN + Nn) \times 10 .$$