PAGE NO. :
DATE : / /

AI61003 Linear Algebra for AI & ML
Assignment 01 - Problem 08

(a) In step 1, we find, for every sample, the "closest" of the k cluster representatives. The distance b/w the sample x_i and the representative z_j is measured by the norm $\|x_i - z_j\|$.
Let us assume that in our application $\|\cdot\| = \|\cdot\|_2$ (2-norm)

Since for every sample x_i and every cluster representative z_j we have to compute $\|x_i - z_j\|_2$ (along with other constant number of constant-time operations), the computational complexity $T_1 = NKt$, where t is the iterations needed to compute 2-norm of a difference of two vectors.

t = (time to compute difference) + (time to compute 2-norm)
$= n + 2n = 3n$

$\Rightarrow T_1 = \boxed{3nkN}$

(b) Let after first step, $|C_i|$ = size of i^{th} cluster = s_i for $i = 1 \ldots k$.
In second step for every cluster we compute the average of all

PAGE NO.:
DATE: / /

the vectors in the cluster.

$t_i = (s_i - 1)n + 1$

= (no. of additions) x (computations for one addition) + (division by s_i)
computational

(t_i is the time reqd. to compute i^{th} cluster representative).

So the computational time complexity

$T_2 = \sum_{i=1}^{k} t_i = n(N-k) + k$

$= nN - (n-1)k$

(use that $\sum_{i=1}^{k} s_i = N$)

✓✓ $\Rightarrow T_2 = nN - (n-1)k$ ($\boxed{\approx nN}$ for large N)

(c) Total computational complexity T of one iteration is $T_{tot} = T_1 + T_2$

$T_{tot} = 3nkN + nN - (n-1)k$

So the no. of computations T_{10} involved in 10 iterations is -

$T_{10} = 10\, T_{tot} = 30nkN + 10nN - 10(n-1)k$

$\approx (30k+10)nN$ (for large N)