HW3 Question 1 Proof

CMSC 423 Fall 2014

Nov. 10 2014

(A) First we show that the MaxDist quantity is non-increasing as more centers are added in farthest-first traversal. Let X^{k-1} be the set of centers selected in the first $k-1$ iterations of farthest-first traversal, and x_k be the center selected in the k-th step, then $\text{MaxDist}(\text{Data}, X) \leq \text{MaxDist}(\text{Data}, X^{k-1})$.

$$
\begin{aligned}
\text{MaxDist}(\text{Data}, X) &= \max_{u \in \text{Data}} \min_{x \in X} d(u, x) && (1)\\
&= \max_{u \in \text{Data}} \min \left[\min_{x \in X^{k-1}} d(u, x), d(u, x_k) \right] && (2)\\
&\leq \max_{u \in \text{Data}} \min_{x \in X^{k-1}} d(u, x) && (3)\\
&= \text{MaxDist}(\text{Data}, X^{k-1}) && (4)
\end{aligned}
$$

From this it follows that $\text{MaxDist}(\text{Data}, X) \leq \text{MaxDist}(\text{Data}, X^t)$ for all t=[1,…,k-1].

(B) Using (A), we show that $d(x_i, x_j) \geq \text{MaxDist}(\text{Data}, X)$ for all $x_i, x_j \in X$.

Assume $i < j$, that is, x_i was chosen as a center in an earlier iteration than x_j. Then,

$$
d(x_i, x_j) \geq \min_{x_t : t=[1,\dots,j-1]} d(x_t, x_j) = \text{MaxDist}(\text{Data}, X^{j-1}) \geq \text{MaxDist}(\text{Data}, X)
$$

where the last inequality follows from (A).

(C) Let $u \in \text{Data}$ be such that $d(u, X) = \text{MaxDist}(\text{Data}, X)$, then by definition $d(u, x) \geq \text{MaxDist}(\text{Data}, X)$ for all $x \in X$.

(D) From (B) and (C), it follows that $X \cup \{u\}$ is a set of $k+1$ points in Data with distance between every pair of points greater than or equal to $\text{MaxDist}(\text{Data}, X)$.

(E) The optimal k clustering must include two of the points defined in (D) in one of it's clusters. Therefore, $\text{MaxDist}(\text{Data}, X_{opt}) \geq \frac{\text{MaxDist}(\text{Data}, X)}{2}$. Which proves the result.