OM Review - XIX

Prof. C. V. Jawahar

IIIT Hyderabad

April 3, 2020

https://forms.gle/o2f5oioWCPRirUgW8

Context for next question

We saw in one of the problems in the last lecture that K-Means is prone to local minima. Consider strategies S1, S2, S3 aimed to solve this issue. Given dataset S we need to find initial set of cluster centroids T.

S1: Pick k points from S that are farthest away from each other.

S2: Pick $x \in S$ uniformly at random and set $T \leftarrow \{x\}$
While $|T| < k$:
(i)Pick $x \in S$ at random, with probabilities directly proportional to $\min_{z \in T} \|x - z\|$
(ii) $T \leftarrow T \cup \{x\}$

S3: Pick $x \in S$ uniformly at random and set $T \leftarrow \{x\}$
While $|T| < k$:
(i)Pick $x \in S$ at random, with probabilities inversely proportional to $\min_{z \in T} \|x - z\|$
(ii) $T \leftarrow T \cup \{x\}$

Question 1

Which of these are true with regards to above?

(A) S2 is the best among S1,S2,S3

(B) S3 is the best among S1,S2,S3

(C) S1 is prone to outliers

(D) S1 is not prone to outliers

More than one may be correct.

Question 2

Which of these is true about following statements on K Means?

(i) Bad initialization can lead to poor convergence speed
(ii) Bad initialization can lead to bad overall clustering
(iii) If we choose $K = |S|$, ie size of dataset; then $J = 0$, hence it is a good choice for K.

(A) (i),(ii),(iii)

(B) None of the others

(C) (i)

(D) (i),(ii)

Question 3

Which of these is true?
(i) MDS with Euclidean distance gives a linear combination of data as the answer
(ii) ISOMAP tries to preserve Euclidean distances between all pairs of points

(A) Both (i) and (ii)

(B) Neither (i) nor (ii)

(C) (i) only

(D) (ii) only

Question 4

Which of these is true about ISOMAP?

(A) It can be obtained by solving for the eigenvectors of $K = -\frac{1}{2}HDH$ where D is the Euclidean distance matrix and H is the centering matrix

(B) It requires us to solve APSP problem once.

(C) It requires us to solve APSP problem more than once.

APSP problem: All pairs shortest paths problem

More than one may be correct.

Question 5

Which of these is true?

(A) It is poosible that LLE can lead to the same solution as PCA

(B) None of the others

(C) In Laplacian Eigenmaps, weights are optimized in one of the steps

(D) In Laplacian Eigenmaps, weights are constants

More than one may be correct