Questions and Answers for K-Means and EM

Bingzhang Zhu

December 15, 2021

1 Why do you prefer Euclidean distance over Manhattan distance in the K means Algorithm?

Euclidean distance is preferred over Manhattan distance since Manhattan distance calculates distance only vertically or horizontally due to which it has dimension restrictions.

On the contrary, Euclidean distance can be used in any space to calculate the distances between the data points. Since in K means algorithm the data points can be present in any dimension, so Euclidean distance is a more suitable option.

2 What are the advantages and disadvantages of the K means Algorithm?

Advantages:
Easy to understand and implement.
Computationally efficient for both training and prediction.
Guaranteed convergence.

Disadvantages:
We need to provide the number of clusters as an input variable to the algorithm.
It is very sensitive to the initialization process.
Good at clustering when we are dealing with spherical cluster shapes, but it will perform poorly when dealing with more complicated shapes.
Due to the leveraging of the euclidean distance function, it is sensitive to outliers.