

Second Part: 15 points

Clustering Analysis (7 points)

1. How can we use k-means to create a hierarchy of clusters?
2. What is the main difference between k-means and k-medoids algorithms?
3. What is the curse of dimensionality and why it is related to clustering?
4. Describe briefly how DB-Scan algorithm works
5. What is the main problem that BIRCH Algorithm tries to solve?

Data Mining in Real World Applications (8 points)

Assume a social network application where users participate to a social platform to share activities and experiences. Users are connected using friendship relationships.

1. Model the social network as a graph specifying its edges and vertices
2. Which algorithm can we use to identify the most authoritative (important) users in the network? Explain briefly the intuition behind.
3. Suppose now that we need to disseminate messages through all the network as fast as possible. Would it be enough to target mainly authoritative users? If the answer is yes, explain how. If the answer is no, which kind of additional information is needed and how it is integrated to find the right users for message dissemination?
4. Each user in the network provides content for which he receives feedbacks from other users in the network. A feedback can be positive or negative. How would you update the graph with this new information?

