Purevmaa Damdinsuren

CPSC 4310/Spring 2020

WR7

**Questions from textbooks and articles:**

1. *There are different types of clustering methods: partitioning methods (e.g., K-means method), density-based methods (e.g., DBSCAN method), probabilistic methods (e.g., Gaussian Mixture Model), hierarchical methods, etc.*

Suppose that you are to allocate a number of automatic teller machines (ATMs) in a given region so as to satisfy a number of constraints. The expected similarity measure in clustering houses to allocate ATMs in a city should be done. Suppose that the clustering is constrained by obstacle objects, such as bridges, rivers, and highways that can affect ATM accessibility. How can a clustering method such as k-means be modified for quality clustering under this constraint?

The constraints given in this case are rivers, bridges, and highways, which I see as kind of border between instances that should not be crossed. Therefore, we can specifically say that instances behind the rivers/highways/bridges should not be belong in as the same cluster as instances that are located front of that rivers/bridges/highways. As I researched that is called Cannot-Link.

1. *What are the main motivations for reducing a dataset's dimensionally? What are the main drawbacks?*

Main motivation is a data visualization because people, we, are not great at visualizing dataset that is more than 3D. It makes it easy to see the common patterns among instances in dataset. It also speeds up the training process. However, it is possible that it can make the model perform slightly worse since it causes some information loss.

1. *Can Principal Component Analysis (PCA) be used to reduce the dimensionality of a highly nonlinear dataset?*

Normally, it can be applied to any dimensional reduction problem of any quantitative data. However, using PCA to reduce the dimension of non-linear dataset can distort the real meaning/difference between instances in dataset since PCA works better if dataset is linearly correlated. In extreme case, result from PCA will be no meaning at all.