Clustering Exercises

Rebecca C. Steorts, Duke University

STA 325

Agenda

- We will consider at some longer exercises as a class on clustering to help prepare for your upcoming exams but also get more practical experience with working with the methods/algorithms that have been introduced in the class. There are three in total, and are highlighted below.
- Specifically, we will go more in depth into hierarchical clustering and mixture models to explore these more thoroughly.

Agenda

As you go through these, please do the following (especially on the mixture model problem):

1. Read through all the parts of the problem.
2. Talk with classmates about how to approach the solution.
3. Utilize class resources or think on your own about how to approach the problem. It might help to write down and summarize something that you're not sure about before proceeding.
4. Do not proceed to for example part II before understanding and make sure that you have completed part I.
5. Be aware that you will work at your own pace and this is okay! Some students may work faster on the theory part or faster on the coding part, etc. The most important part is that you work at your own pace and have a full understanding of the problem.

Exercise: Mixture Models

There are two exercises with solutions on mixture models that can be found here:

https://github.com/resteorts/data-clean/blob/main/exercises/exercise-intro-to-mixture-models.pdf

Exercise: Application to Tennis data set

The full exercise can be found here:

https://github.com/resteorts/data-clean/tree/main/exercises/exericise-clustering-tennis

Please work on this on Monday, November 4 in OH if you need assistance and help before solutions are posted. This is a non-graded exercise/homework assignment.

Exercise: Exponential Mixture Models

The exercise on exponential mixture models can be found here:

https://github.com/resteorts/data-clean/blob/main/exercises/exercise-mixture-models/exercise-exponential.pdf

The tutorial on mixture models can be found here:
https://arxiv.org/pdf/1901.06708

Solutions on mixture models will be updated here:

https://github.com/resteorts/data-clean/tree/main/exercises/exercise-mixture-models/solutions/