Clustering Methods

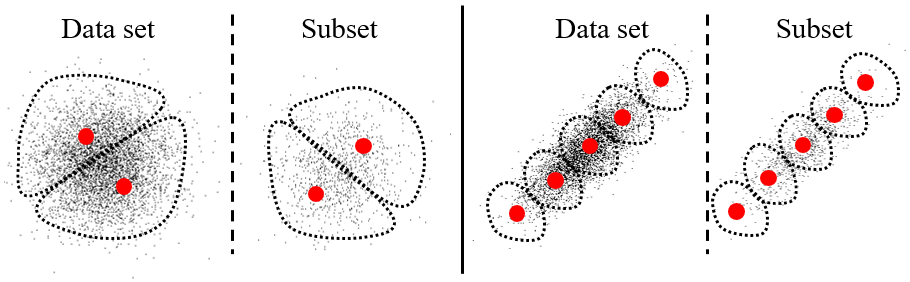
Exercises 3/7

1. Branch-and-bound algorithm uses merges to generate all possible clusterings. Define similar algorithm based on cluster splits which starts by having all datapoints in a single cluster and then split the cluster. Is it possible to define similar bounding criterion as we had with the merge-based algorithm? If yes, show how. If not, why not?

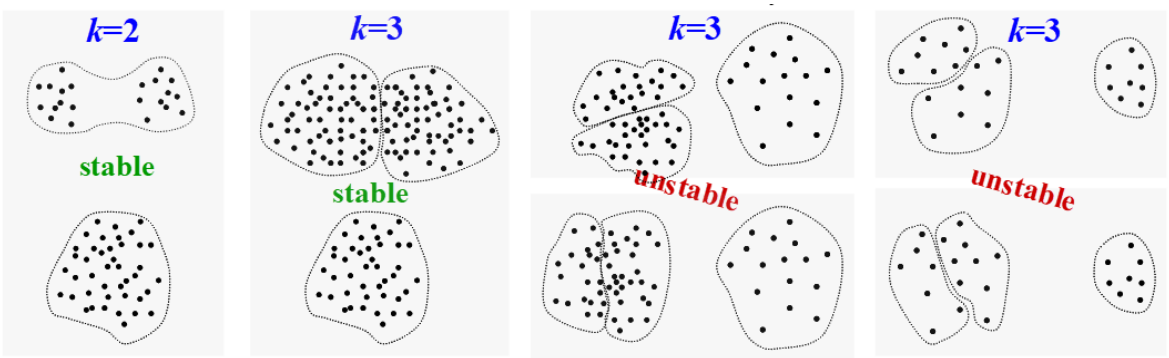
2. Invent an alternative approach to create all possible clusterings without merging or splitting.

3. What practical use do we have if we know how stable clustering is? Can you give one or few examples of use scenarios?

4. Stability has also been used for detecting the number of clusters[1] by measuring the stability of clustering with different k values. The hypothesis is that clustering is unstable with wrong choice of k. Below are two examples. How do you interpret the results? Do they support the hypothesis? If not, what is the root problem?



5. Four more examples and same questions as in task 4.



[1] **Rezaei and Fränti**, Can the number of clusters be determined by external indices? *IEEE Access*, 2020.