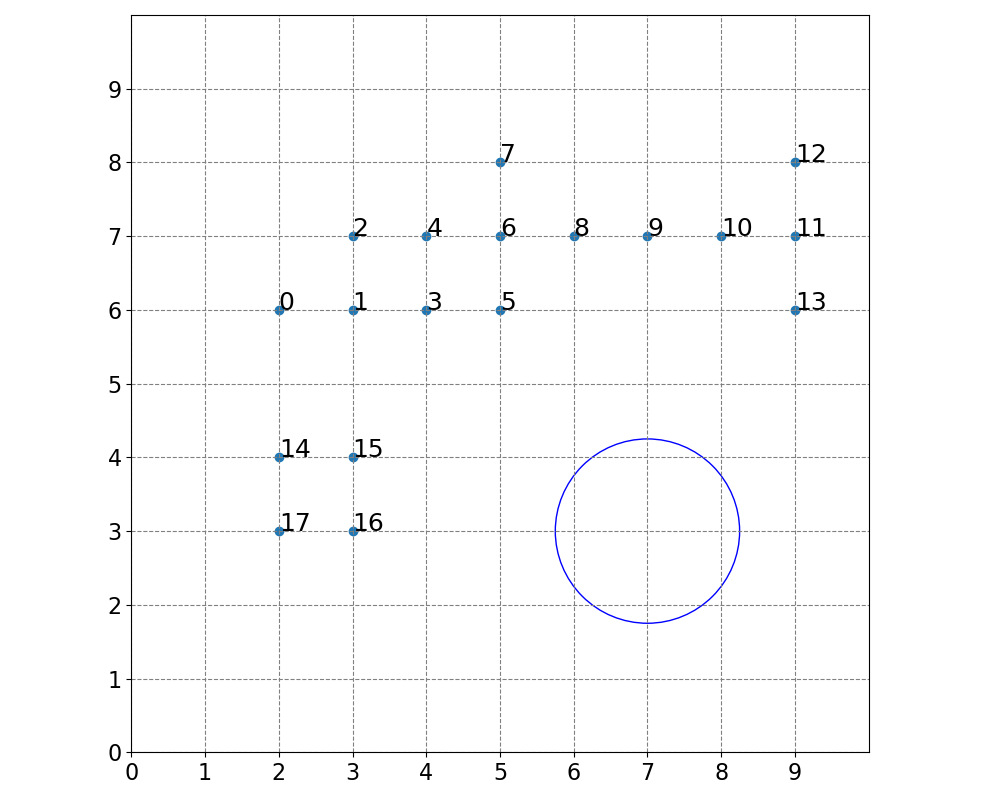
**CS5228 Tutorial 2 - Clustering**

**Q1: DBSCAN**

The following figure shows a toy dataset with 17 points.



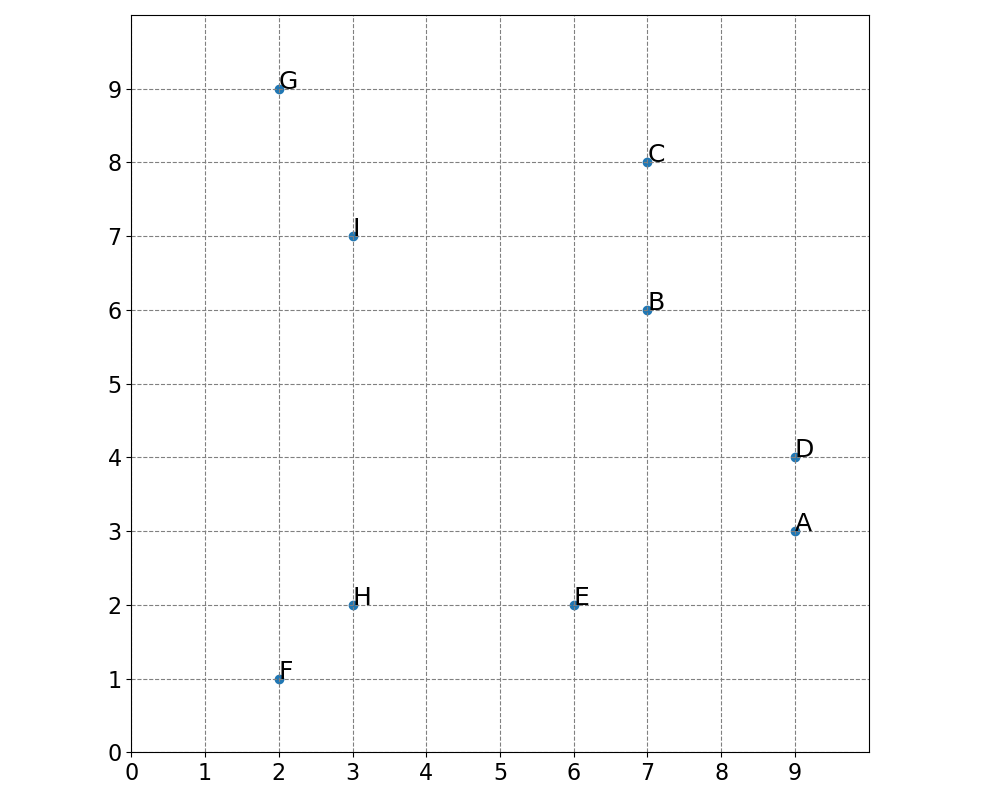
1. Run DBSCAN by hand with and . (The blue circle shows a radius of ) List all core points, border points, and outliers (noise points).
2. How many clusters are there, and what are their data points?
3. Can you add 2 data points such that the resulting clustering contains only 1 cluster and no noise?

**Q2: DBSCAN vs K-Means**

1. Name a few fundamental differences between K-Means and DBSCAN.
2. Name a few meaningful criteria to decide whether to use K-Means or DBSCAN on a task?
3. Name a few example tasks and discuss whether K-Means or DBSCAN would be your method of choice.

**Q3: Hierarchical Clustering**

Compared to K-Means and DBSCAN, hierarchical clustering (specifically Agglomerative Nesting or AGNES) is a hierarchical clustering method. As such, each data point may belong to different clusters depending on the hierarchy level (e.g. higher level clusters can contain lower level clusters). AGNES yields complete clusterings, where each point belongs to at least one cluster.



1. Perform AGNES by hand with single linkage. Write down the steps involved (in any format).

­­­ **Q4: Clustering Evaluation**

There is no perfect way to evaluate clusters – though using “ground truth” or external cluster evaluation is generally preferred if our ground truth label is indeed meaningful and of interest. Overall, it is important to be aware of the limitations and interpretation of different evaluation methods.

1. What are some limitations of using With Cluster Sum of Squares (WCSS) as a cluster evaluation metric?

1. Which limitations of WCSS does the Silhouette Score successfully mitigate?