MA477: Data Science

Lesson 34 Board Sheet — 21 April 2026

United States Military Academy, West Point
Instructor: MAJ Patrick Kuiper

---

1 Clustering Lesson Objectives

- Understand the k-means and hierarchical clustering algorithms.
- Use functions in sklearn to perform clustering, such as DBSCAN.
- Understand practical issues in clustering.

2 Hierarchical Clustering and K-Means Problems

2.1 Problem 1: Bottom-Up Hierarchical Clustering

Consider the following three data points, each with two features:

$$A = (1, 1), \qquad B = (2, 2), \qquad C = (5, 5).$$

Use **agglomerative hierarchical clustering** with **Euclidean distance** and **single linkage**. Compute the pairwise distances and show the order in which the clusters merge.

2.2 Problem 2: K-Means Clustering

For this problem, interpret the four points as

$$A = (1, 1), \qquad B = (1, 3), \qquad C = (9, 4), \qquad D = (10, 11).$$

Use **k-means clustering** with $K = 2$. Suppose the initial cluster centers are chosen randomly as points B and A.

> Compute the distance from each point to each initial cluster center. Assign each point to the nearest center. Then compute the updated cluster centers after the first iteration.