

BMEG3105 Lecture 8 scribing (Ma Wing Laam)

- Clustering Analysis: Finding groups of objects such that the objects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups.
- Intra-cluster differences vs Inter-cluster differences
- Hierarchical clustering: To compute a similarity or distance matrix for hierarchical clustering, start by calculating the initial distances between all data points, treat each point as its own cluster, then iteratively merge the two closest clusters, update the distance matrix accordingly, and repeat this process until only a single cluster remains.
- Distance metrics: Cosine similarity, Correlation, Euclidean distance, Manhattan distance, Mahalanobis distance

- Euclidean distance:

$$\text{Distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

- Correlation:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

- Mahalanobis distance: unlike Euclidean distance, it only consider straight line.