Clustering and Dimensionality Reduction

Swati Mishra

Applications of Machine Learning (4AL3)

Fall 2024

McMaster University

ENGINEERING

Review

- Generative Models: Clustering
- K- Means Clustering
- Training and Evaluations
- Brief stint with Unsupervised Learning

Clustering… contd from last lecture

- Types of Clustering:
  - K-Means

pre-specified clusters

1. Randomly assign a number, from 1 to $\boxed{K}$, to each of the observations. These serve as initial cluster assignments for the observations.

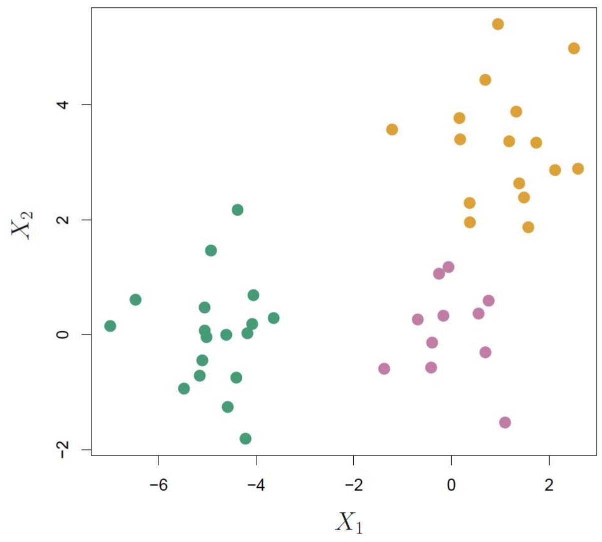
2. Iterate until the cluster assignments stop changing:

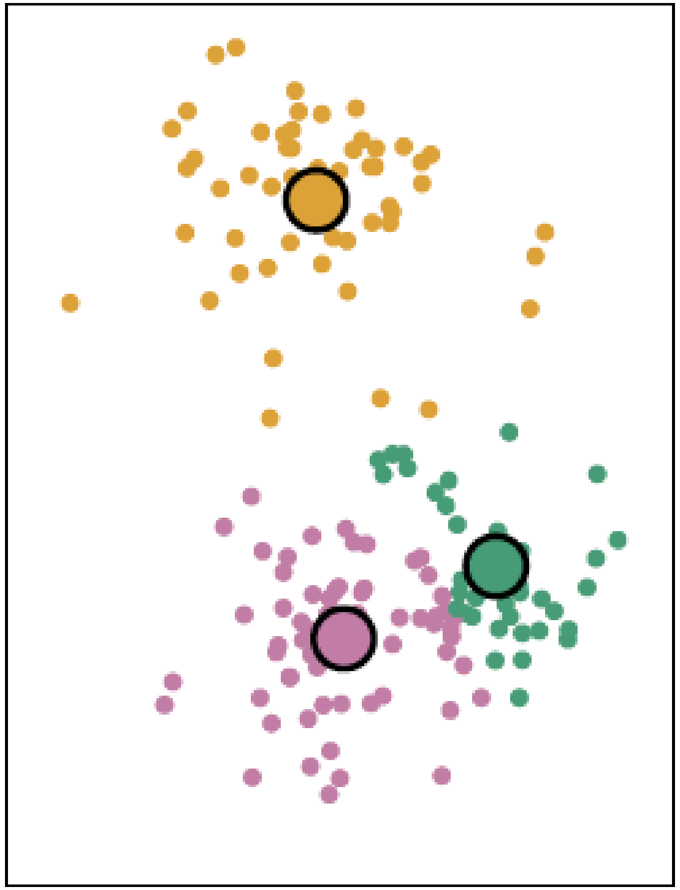
   (a) For each of the K clusters, compute the cluster *centroid*. The kth cluster centroid is the vector of the p feature means for the observations in the kth cluster.

   (b) Assign each observation to the cluster whose centroid is closest (where *closest* is defined using Euclidean distance).

Clustering

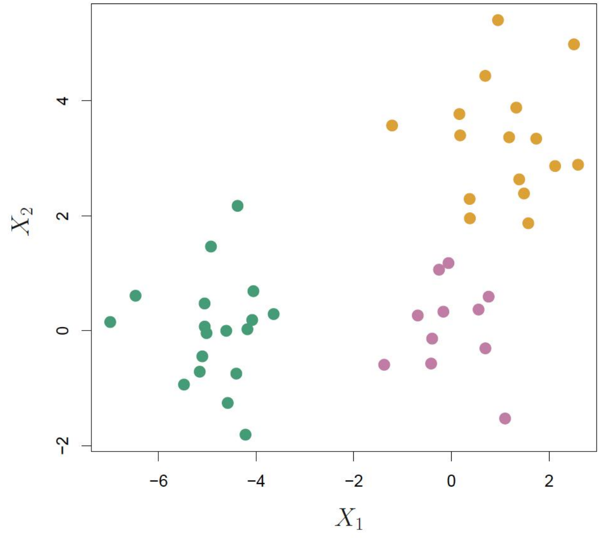
- Types of Clustering:
  - K-Means

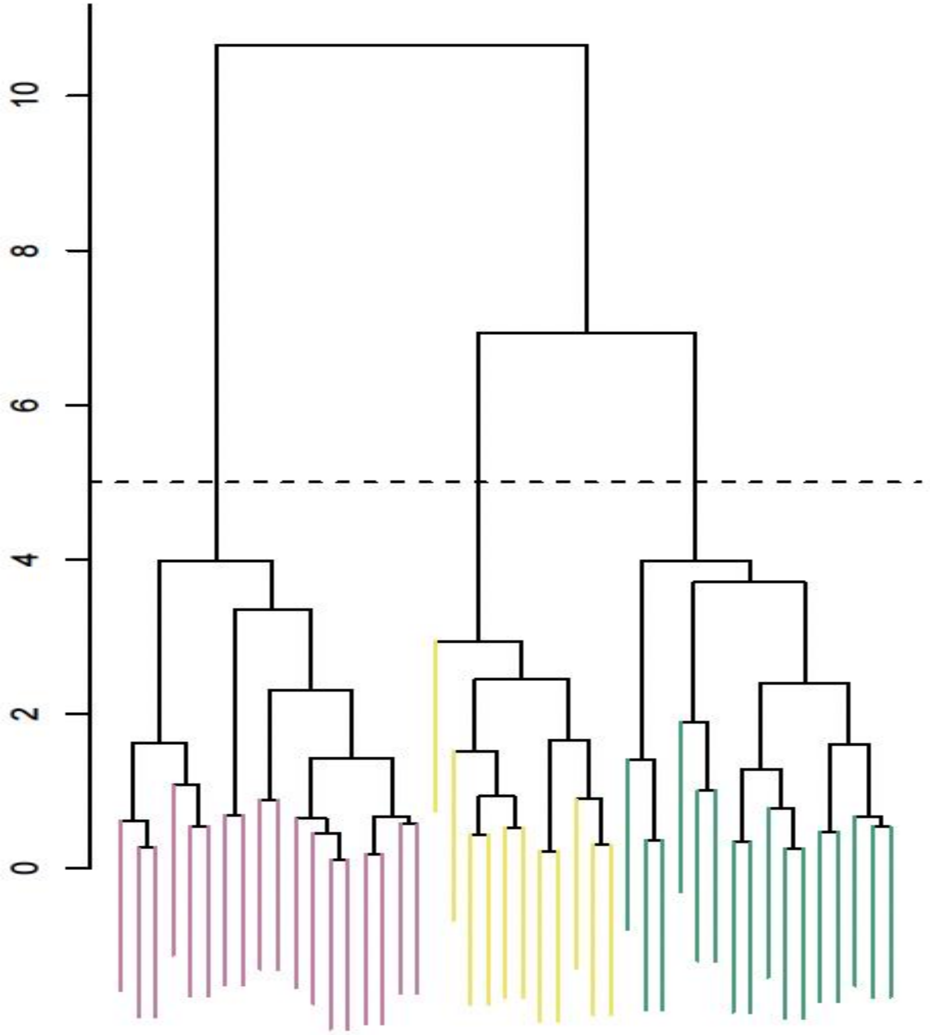




Clustering

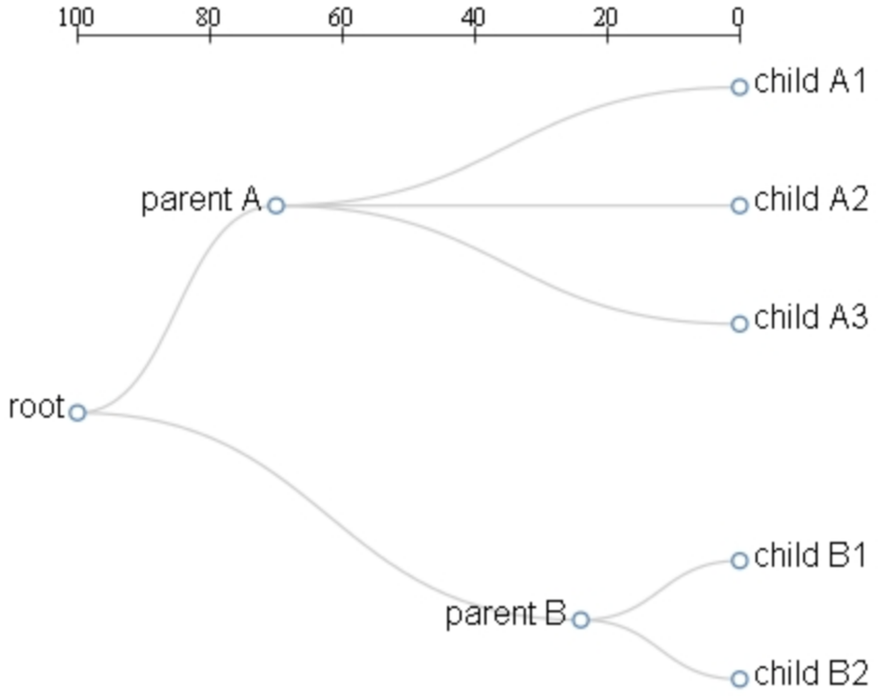
- Types of Clustering:
  - K-Means
  - Hierarchical clustering





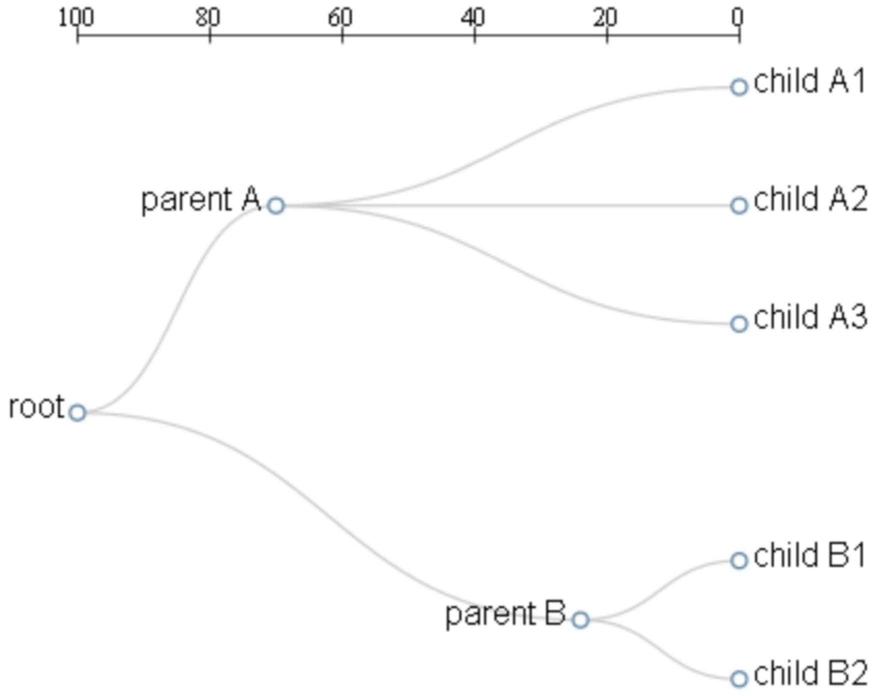
Clustering

- Types of Clustering:
  - K-Means
  - Hierarchical clustering
    - Uses tree-based representation of observations called Dendrograms.



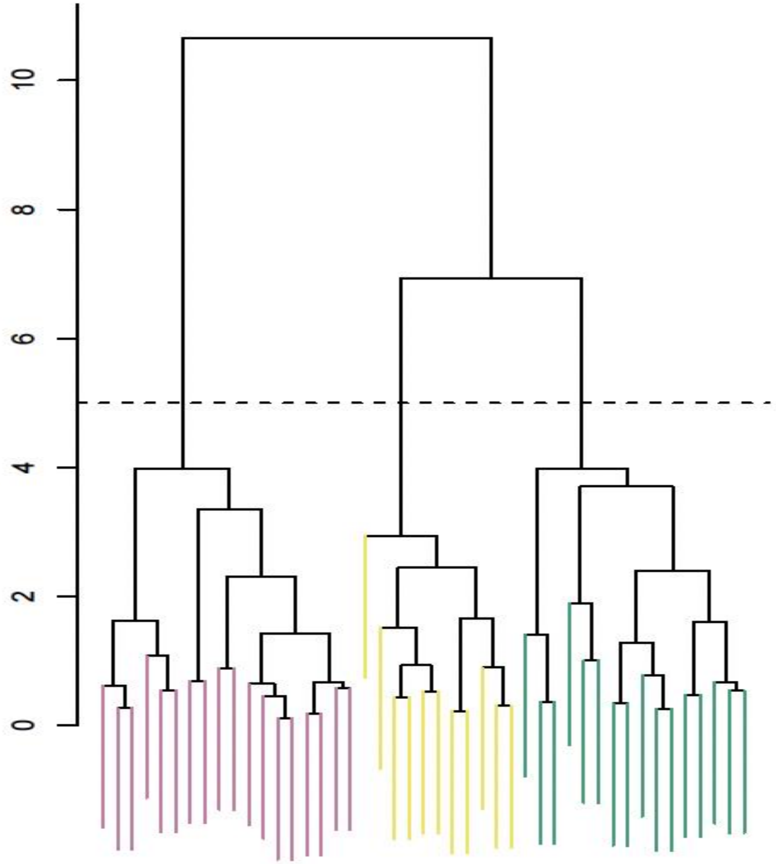
Clustering

- Types of Clustering:
  - K-Means
  - Hierarchical clustering
    - Uses tree-based representation of observations called Dendrograms.
    - The lower in the tree, fusions occur, the more similar the groups of observations are to each other.



Clustering

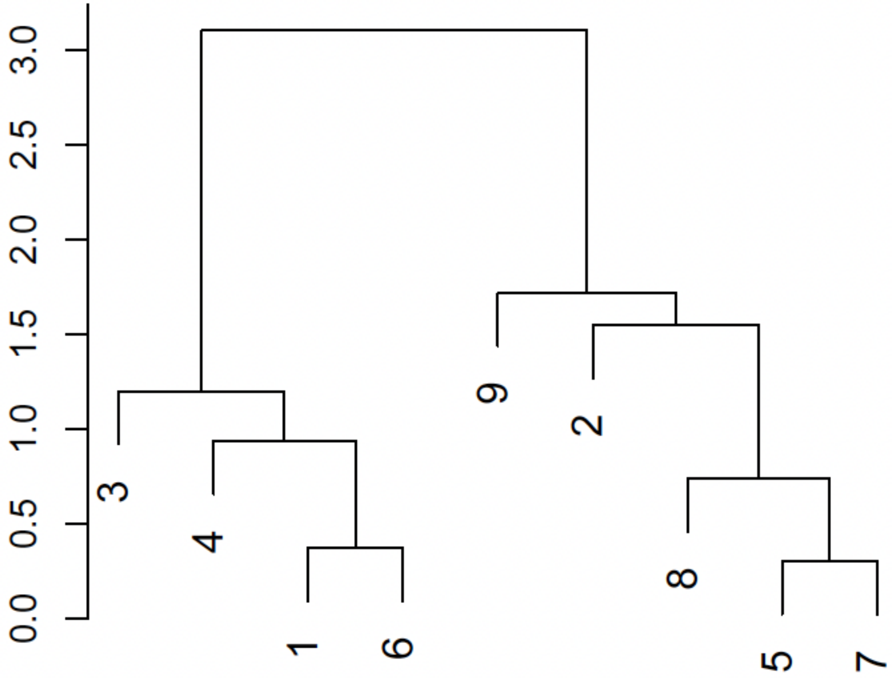
- Types of Clustering:
  - K-Means
  - Hierarchical clustering
    - Uses tree-based representation of observations called Dendrograms.
    - The lower in the tree, fusions occur, the more similar the groups of observations are to each other.
    - Most common type is *bottom-up* or *agglomerative*



Interpreting Dendrograms

- Types of Clustering:
  - K-Means
  - Hierarchical clustering

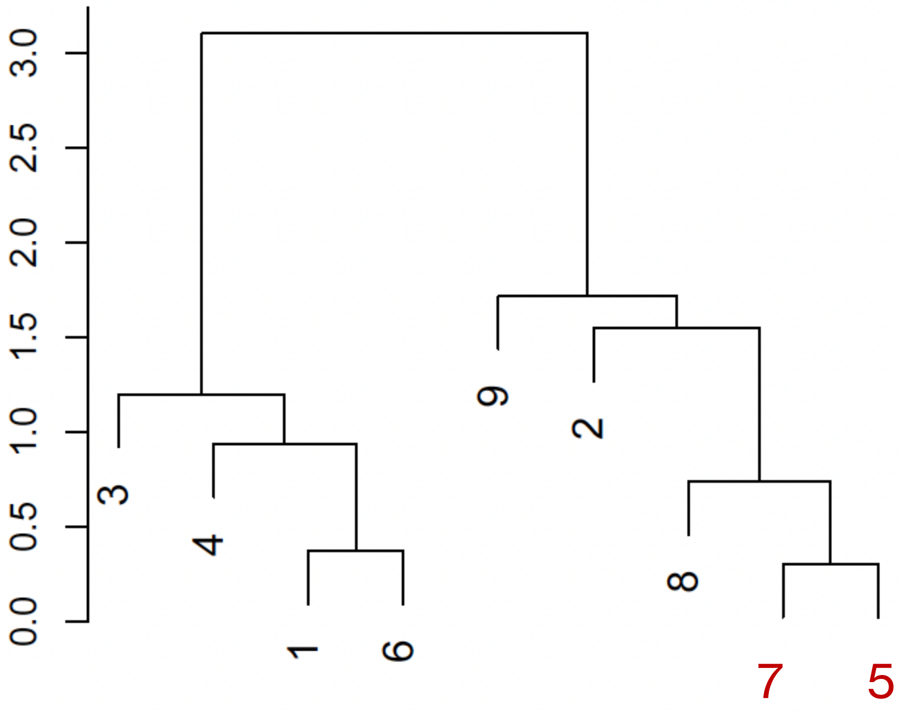
Which observations are similar?



Interpreting Dendrograms

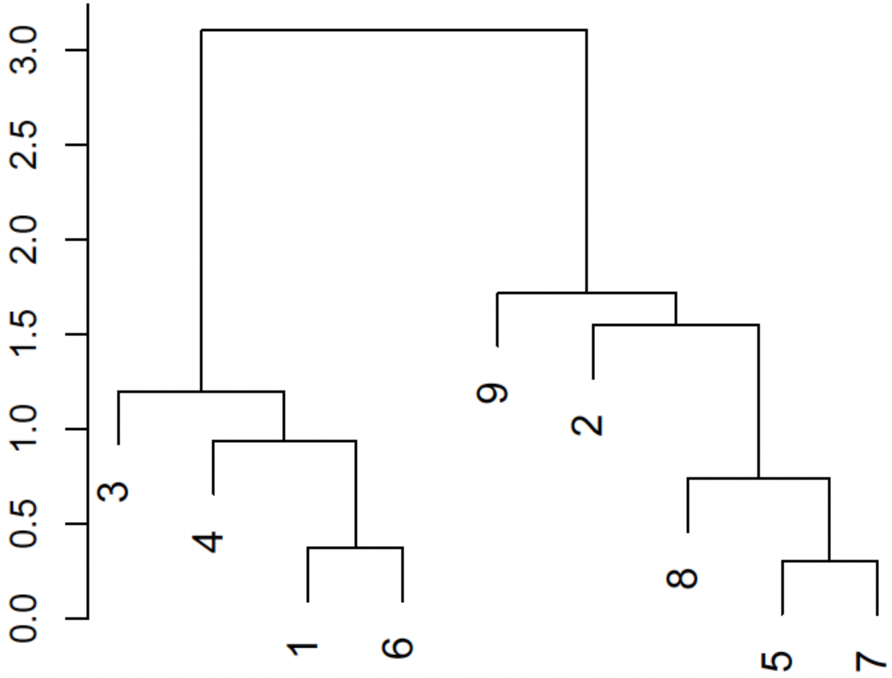
- Types of Clustering:
  - K-Means
  - Hierarchical clustering

What happens when we change the order?



Interpreting Dendrograms

- Types of Clustering:
  - K-Means
  - Hierarchical clustering
    - For n observations, 2^{n-1} reordering of dendrograms.



Which observations are similar?

Hierarchical Clustering

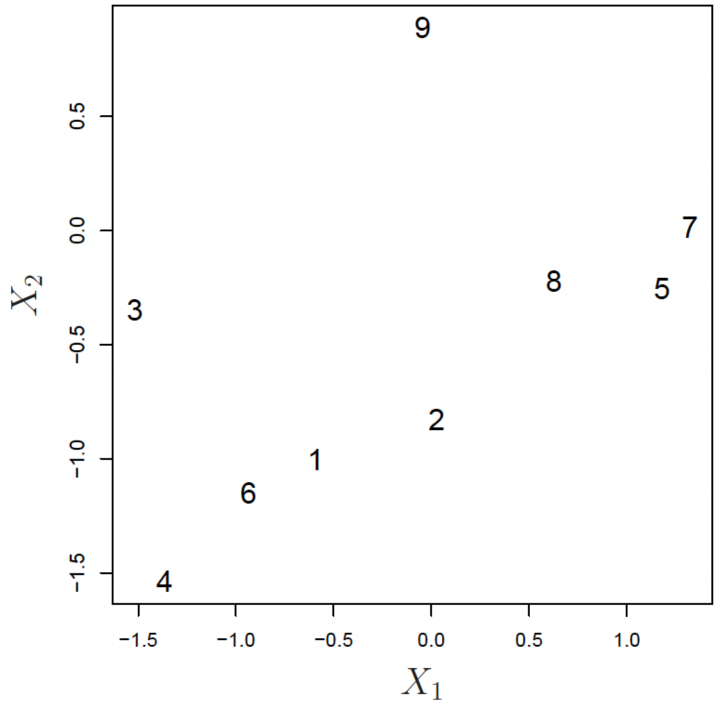
1. Begin with n observations and a measure (such as Euclidean distance) of all the $\binom{n}{2} = n(n-1)/2$ pairwise dissimilarities. Treat each observation as its own cluster.

2. For $i = n, n-1, \ldots, 2$:

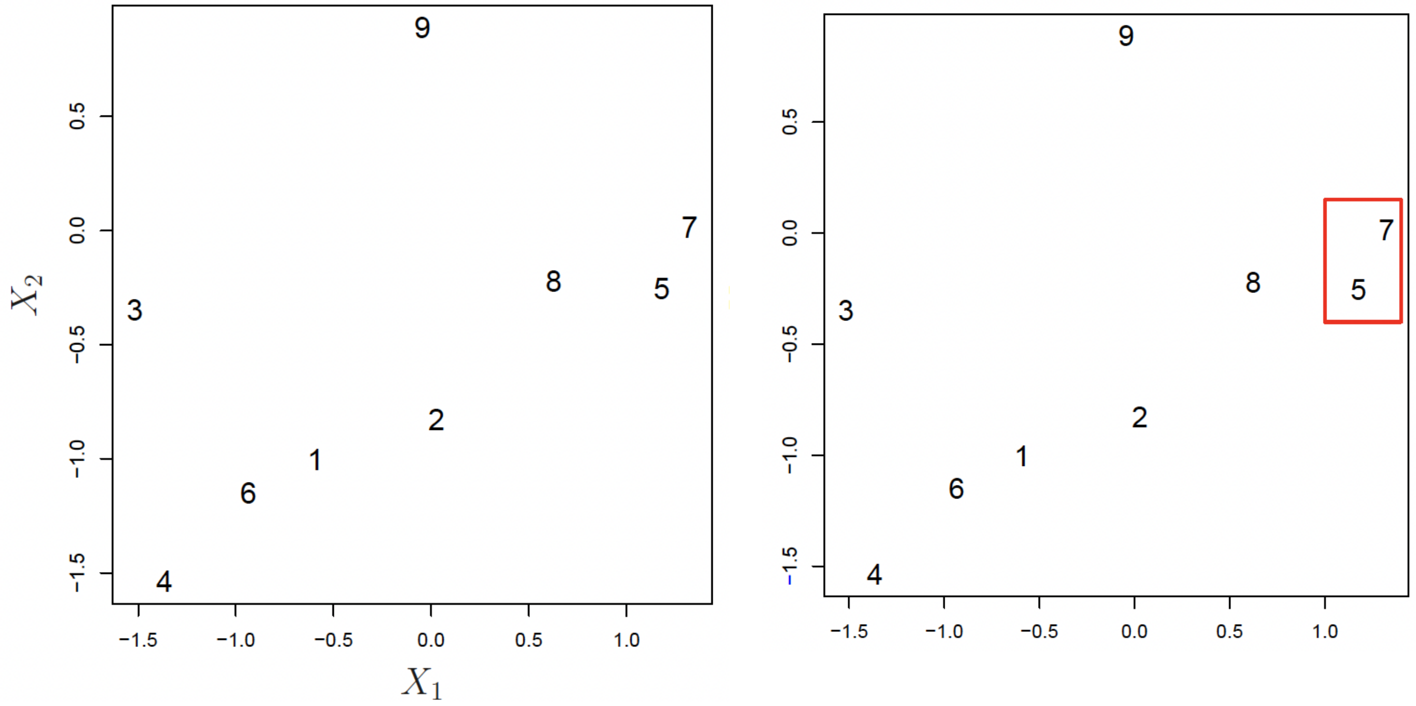
   (a) Examine all pairwise inter-cluster dissimilarities among the i clusters and identify the pair of clusters that are least dissimilar (that is, most similar). Fuse these two clusters. The dissimilarity between these two clusters indicates the height in the dendrogram at which the fusion should be placed.

   (b) Compute the new pairwise inter-cluster dissimilarities among the $i - 1$ remaining clusters.

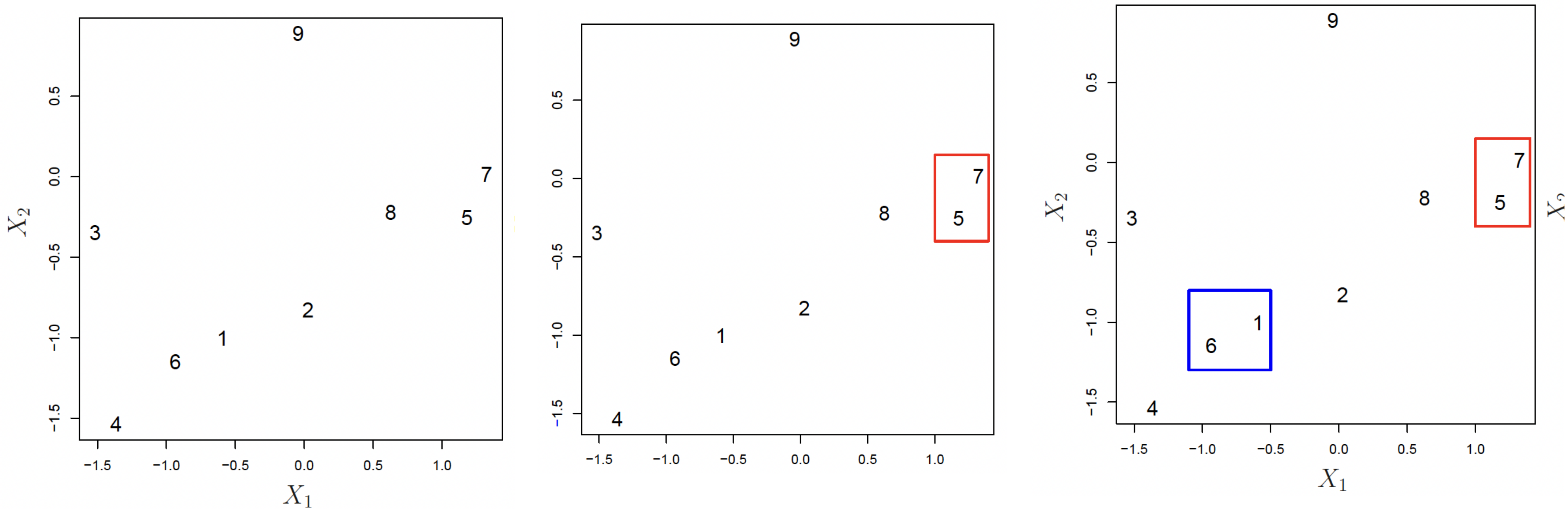
Hierarchical Clustering



Hierarchical Clustering

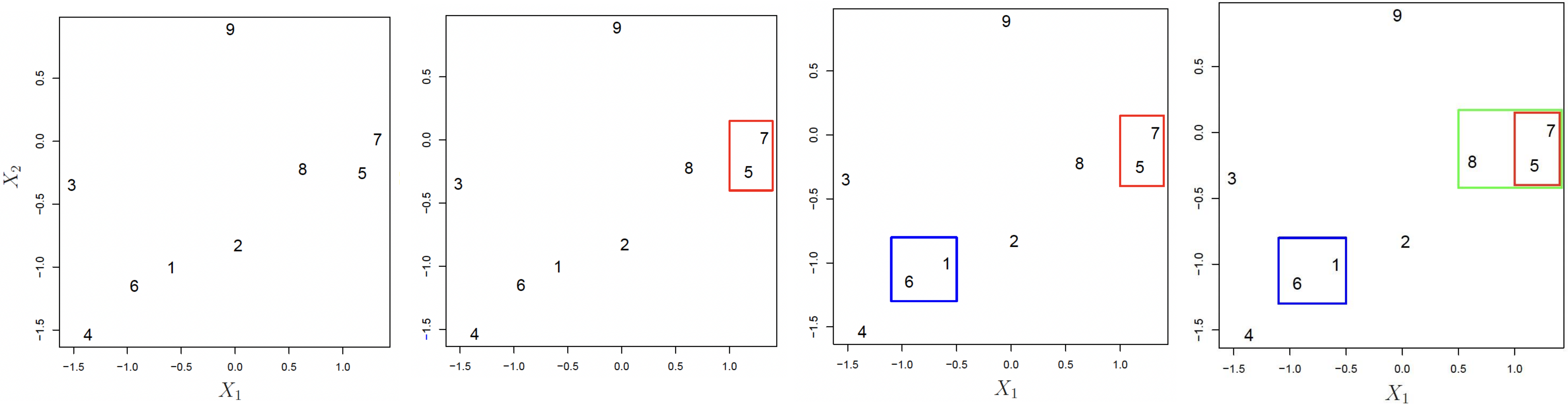


Hierarchical Clustering



McMaster University

Hierarchical Clustering



Hierarchical Clustering

- Design considerations:
  - Linkage Techniques.
    - **Complete:** Maximal intercluster dissimilarity.
    - **Single:** Minimal intercluster dissimilarity.
    - **Average**: Mean intercluster dissimilarity
    - **Centroid**: Dissimilarity between the centroid

Hierarchical Clustering

- Design considerations:
  - Linkage Techniques.
    - **Complete:** Maximal intercluster dissimilarity.
    - **Single:** Minimal intercluster dissimilarity.
    - **Average**: Mean intercluster dissimilarity
    - **Centroid**: Dissimilarity between the centroid

Hierarchical Clustering

- Design considerations:
  - Linkage Techniques.
    - **Complete:** Maximal intercluster dissimilarity.
    - **Single:** Minimal intercluster dissimilarity.
    - **Average:** Mean intercluster dissimilarity
    - **Centroid**: Dissimilarity between the centroid

Hierarchical Clustering

- Design considerations:
  - Linkage Techniques:
    - **Complete:** Maximal intercluster dissimilarity.
    - **Single:** Minimal intercluster dissimilarity.
    - **Average:** Mean intercluster dissimilarity
    - **Centroid**: Dissimilarity between the centroid

Hierarchical Clustering

- Design considerations:
  - Linkage Techniques:
    - **Complete:** Maximal intercluster dissimilarity.
    - **Single:** Minimal intercluster dissimilarity.
    - **Average:** Mean intercluster dissimilarity
    - **Centroid**: Dissimilarity between the centroid.
  - Measures:
    - Euclidean Distance
    - Correlation based measures

Hierarchical Clustering

- Design considerations:

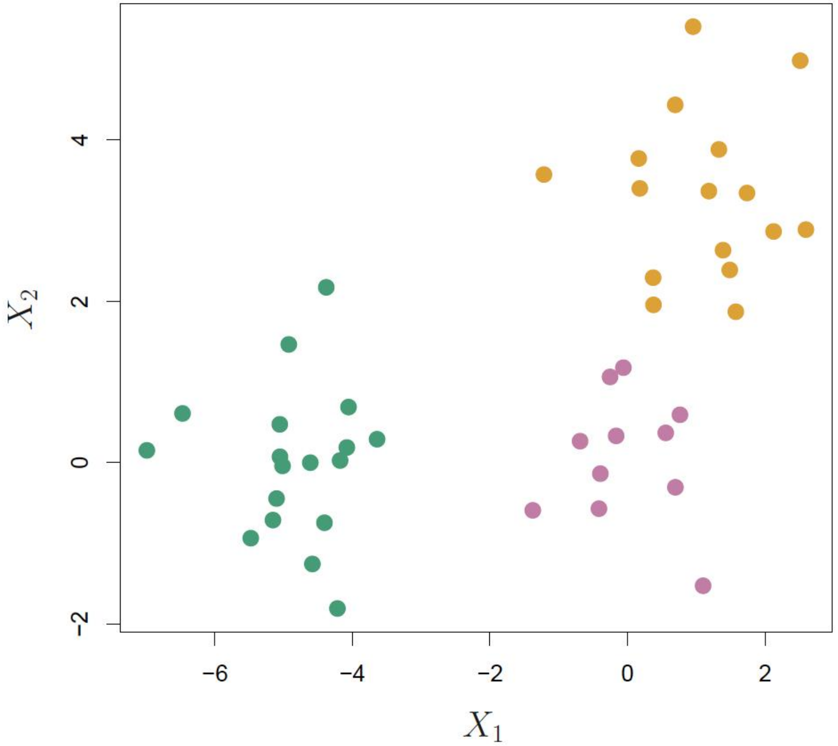
1. Begin with n observations and a measure (such as Euclidean distance) of all the $\binom{n}{2} = n(n-1)/2$ pairwise dissimilarities. Treat each observation as its own cluster.

2. For $i = n, n-1, \ldots, 2$:

   (a) Examine all pairwise inter-cluster dissimilarities among the i clusters and identify the pair of clusters that are least dissimilar (that is, most similar). Fuse these two clusters. The dissimilarity between these two clusters indicates the height in the dendrogram at which the fusion should be placed.

   (b) Compute the new pairwise inter-cluster dissimilarities among the $i-1$ remaining clusters.

Projections

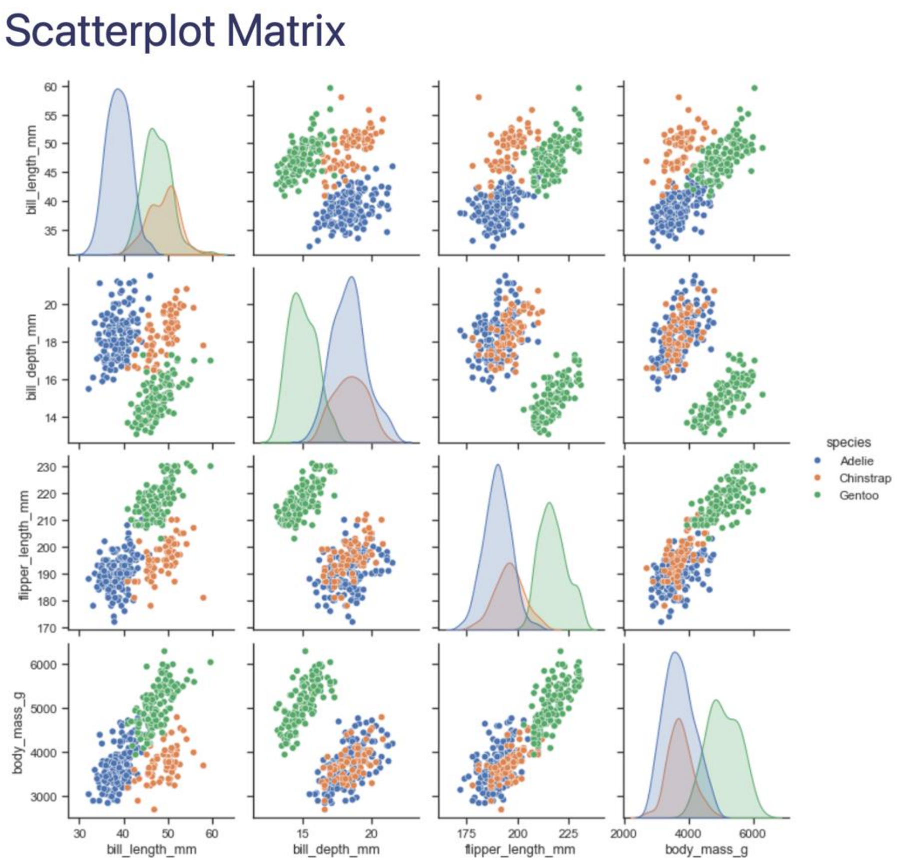


How do we project a samples in this space?

Principal Component Analysis

- Suppose that we wish to visualize n observations with measurements on a set of p features.

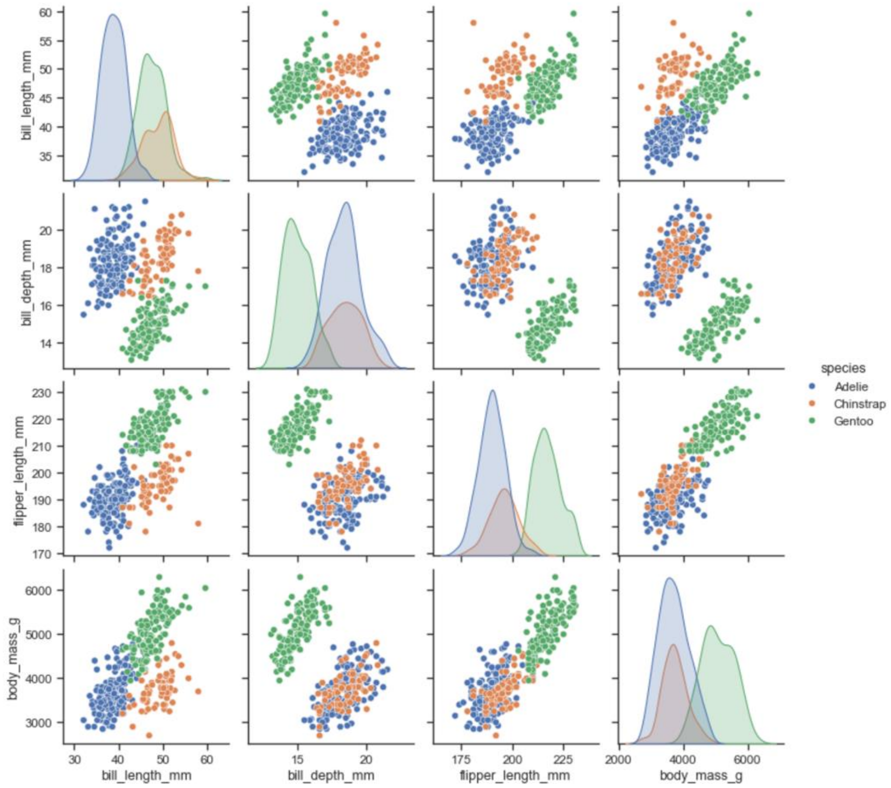
- Not all of p dimensions are equally interesting.



Principal Component Analysis

- Suppose that we wish to visualize n observations with measurements on a set of p features.
- Not all of p dimensions are equally interesting.
- PCA finds a low-dimensional representation of a data set that contains as much as possible of the variation.

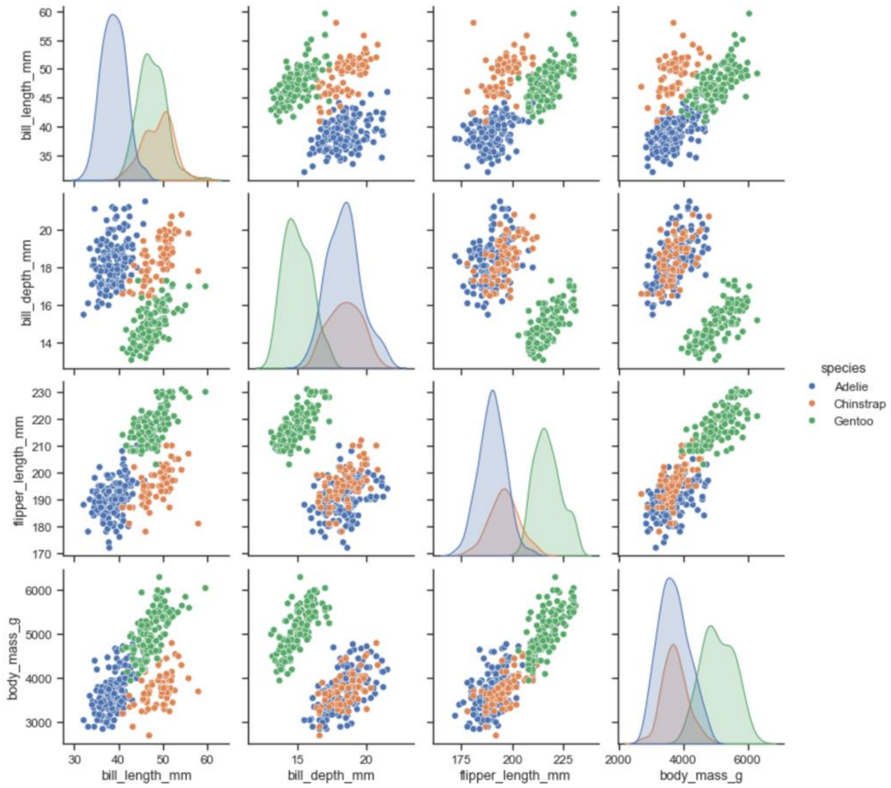
Scatterplot Matrix



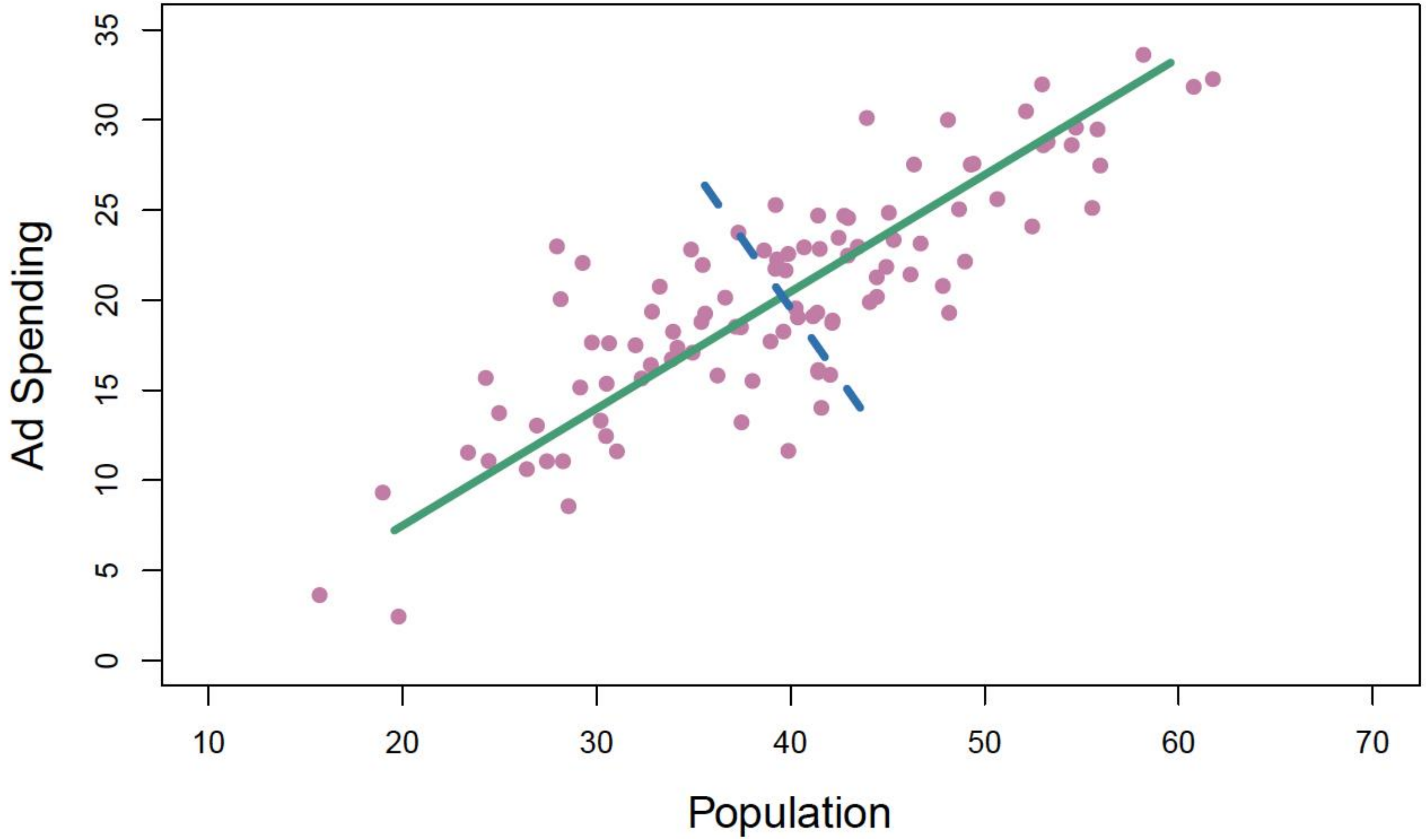
Principal Component Analysis

- Suppose that we wish to visualize n observations with measurements on a set of p features.
- Not all of p dimensions are equally interesting.
- PCA finds a low-dimensional representation of a data set that contains as much as possible of the variation.
- PCA seeks a small number of dimensions that are as interesting as possible.

Scatterplot Matrix



Principal Component Analysis



Principal Component Analysis

- First Principal Component:
  - Normalized linear combination of the features that has the largest variance.

$$Z = \emptyset_{11}X_1 + \emptyset_{21}X_2 + \ldots + \emptyset_{p1}X_p$$

  - $\emptyset_{11} =$ loadings of the PCA

Principal Component Analysis

- First Principal Component:
  - Normalized linear combination of the features that has the largest variance.

$$Z = \emptyset_{11}X_1 + \emptyset_{21}X_2 + \ldots + \emptyset_{p1}X_p$$

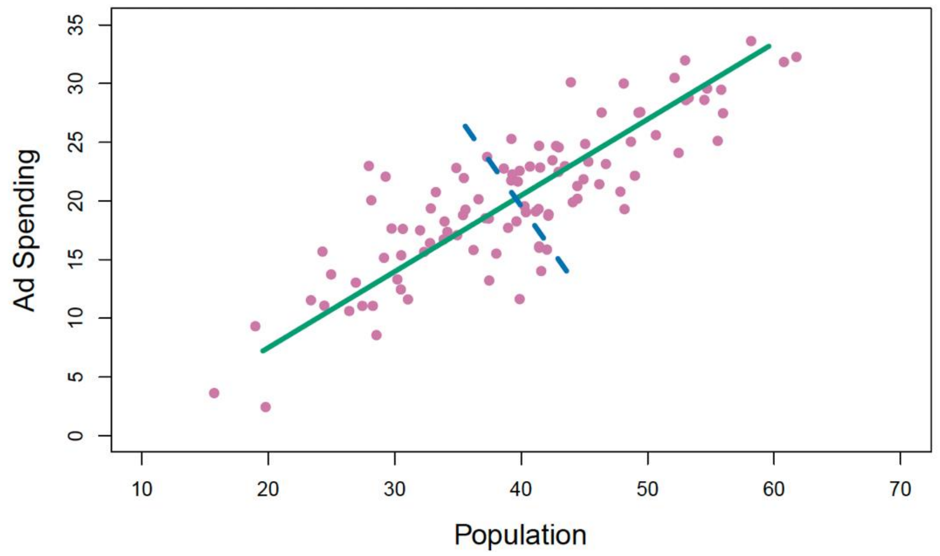
- $\emptyset_{11}$ = loadings of the PCA → define a direction in feature space along with the data varies the most

Principal Component Analysis

- First Principal Component:
  - Normalized linear combination of the features that has the largest variance.

$$Z = \emptyset_{11}X_1 + \emptyset_{21}X_2 + \ldots + \emptyset_{p1}X_p$$

  - $\emptyset_{11}$ = loadings of the PCA



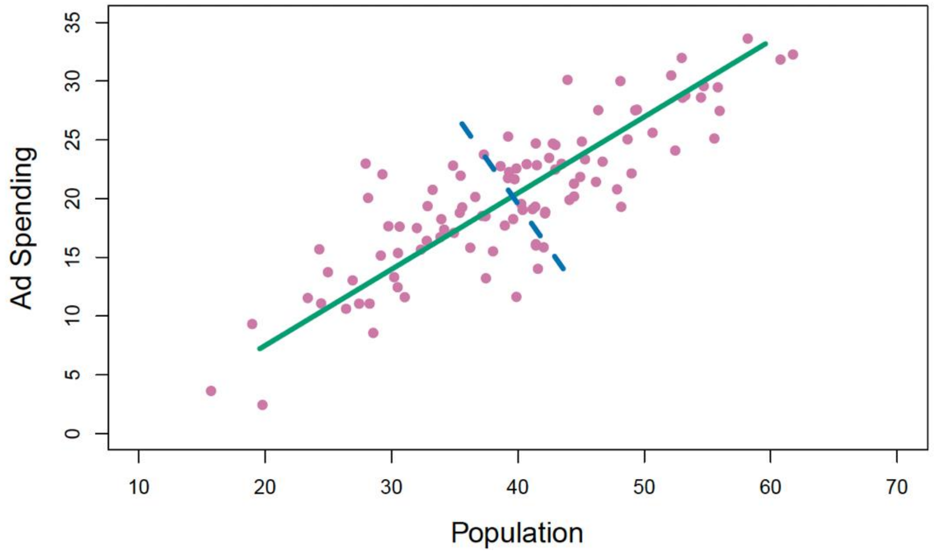
Principal Component Analysis

- First Principal Component:
  - Normalized linear combination of the features that has the largest variance.

$$Z = \emptyset_{11}X_1 + \emptyset_{21}X_2 + \ldots + \emptyset_{p1}X_p$$

- $\emptyset_{11}$ = loadings of the PCA
- Normalized linear combination of the features.

$$\sum_{j=1}^{p} \phi_{j1}^2 = 1$$



Principal Component Analysis

- First Principal Component:
  - Normalized linear combination of the features that has the largest variance.

$$Z = \emptyset_{11} X_1 + \emptyset_{21} X_2 + \ldots + \emptyset_{p1} X_p$$

  - $\emptyset_{11}$ = loadings of the PCA
  - Normalized linear combination of the features $\sum_{j=1}^{p} \phi_{j1}^2 = 1$

How do we compute principal components?

Next Lecture

How do we compute principal components?

Readings

Required Readings:

Introduction to Statistical Learning

- Chapter 12 – Section 12.4 page 526 - 535

Thank You

McMaster University