THE UNIVERSITY OF TENNESSEE KNOXVILLE

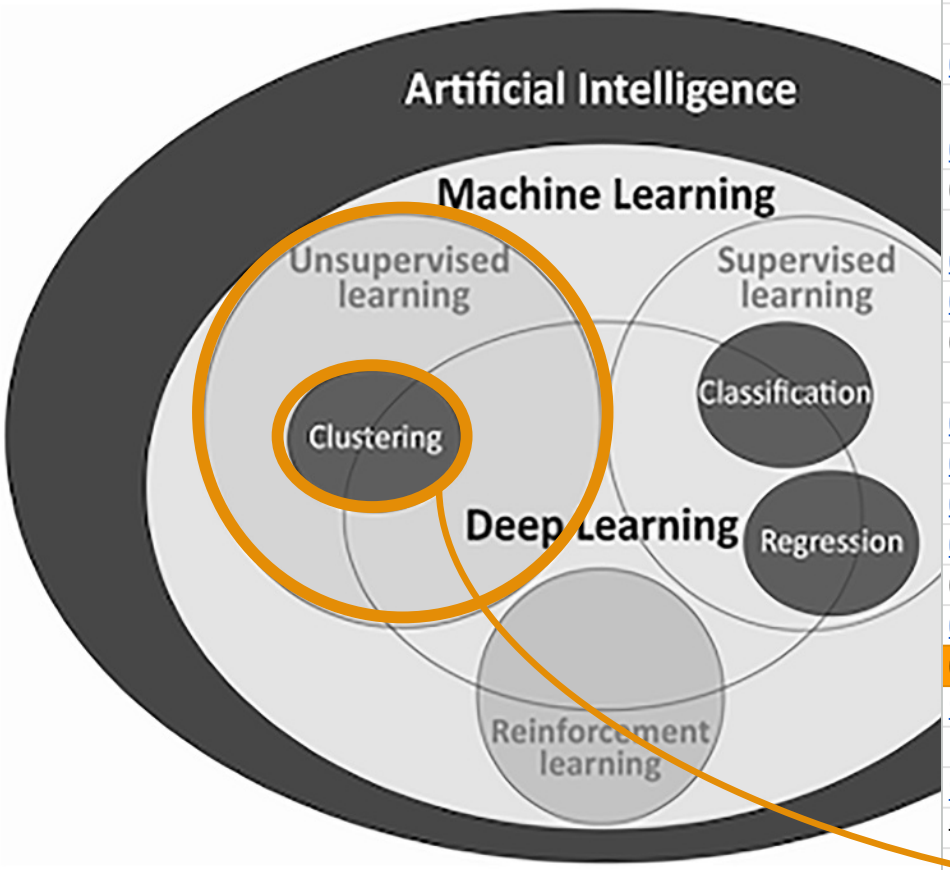
AICIP RESEARCH

COSC 522 – Machine Learning

Unsupervised Learning (Clustering)

Hairong Qi, Gonzalez Family Professor
Electrical Engineering and Computer Science
University of Tennessee, Knoxville
https://www.eecs.utk.edu/people/hairong-qi/
Email: hqi@utk.edu

AICIP RESEARCH



	Part 1: Statistical Methods	
	Baysian Learning	
08/20 (T)		Introduction
08/22 (R)		Baysian Decision Theory and Parametric Learning
08/27 (T)		Baysian Decision Theory and Non-Parametric Learning
08/29 (R)		Case Study: Representation for Natural Language (taught by Andre Cozma)
09/03 (T)		Parametric vs. Non-Parametric Learning: Some In-Depth Discussio
09/05 (R)		Homework and Project Discussion (taught by Fanqi Wang)
	Neural Networks	
09/10 (T)		Biological Neuron and Perceptron
09/12 (R)		Perceptron
09/17 (T)		Back Propagation and Gradient Descent
09/19 (R)		Back Propagation
09/20 (F)		TRUST-AI Seminar
09/24 (T)		Kernel Methods and Review
09/26 (R)	Test 1	
10/01 (T)		Kernel Methods and Support Vector Machine
	Regression	
10/03 (R)		Regression
~~10/08 (T)~~	~~Fall Break (No Class)~~	
	Unsupervised Learning	
10/10 (R)		Logistic Regression; k-means
10/15 (T)		Hierarchical methods and auto-encoder
10/17 (R)		recap

Questions

AICIP RESEARCH

- What is unsupervised learning? What is/are unknown?
- What are the new distance metrics introduced to measure point to cluster distance and cluster to cluster distance?

- What is kmeans? Objective function, optimal solution, procedure, geometrical interpretation
- What is winner-takes-all? Objective function, optimal solution, procedure, geometrical interpretation
- What are the differences between batch processing and online processing? Pros and cons?
- What are the potential issues with kmeans or wta?

- What is hierarchical clustering? What are bottom-up vs. top-bottom hierarchical clustering?
- What is agglomerative clustering and the different linkage of agglomerative clustering
- What is a dendrogram?

Unsupervised Learning

AICIP RESEARCH

- What's unknown?
  - In the training set, which class does each sample belong to?
  - For the problem in general, how many classes (clusters) is appropriate?

Distance from a Point to a Cluster

- Euclidean distance $d_{euc}(x, A) = \|x - \mu_A\|$
- City block distance
- Squared Mahalanobis distance

$$d_{mah}(x, A) = (x - \mu_A)^T \Sigma_A^{-1} (x - \mu_A)$$

Distance between Clusters

AICIP RESEARCH

- The centroid distance

$$d_{mean}(A,B) = \|\mu_A - \mu_B\|$$

- Nearest neighbor measure

$$d_{\min}(A,B) = \min_{a,b} d_{euc}(a,b) \quad \text{for } a \in A, b \in B$$

- Furthest neighbor measure

$$d_{\max}(A,B) = \max_{a,b} d_{euc}(a,b) \quad \text{for } a \in A, b \in B$$

PART I: KNOWING K

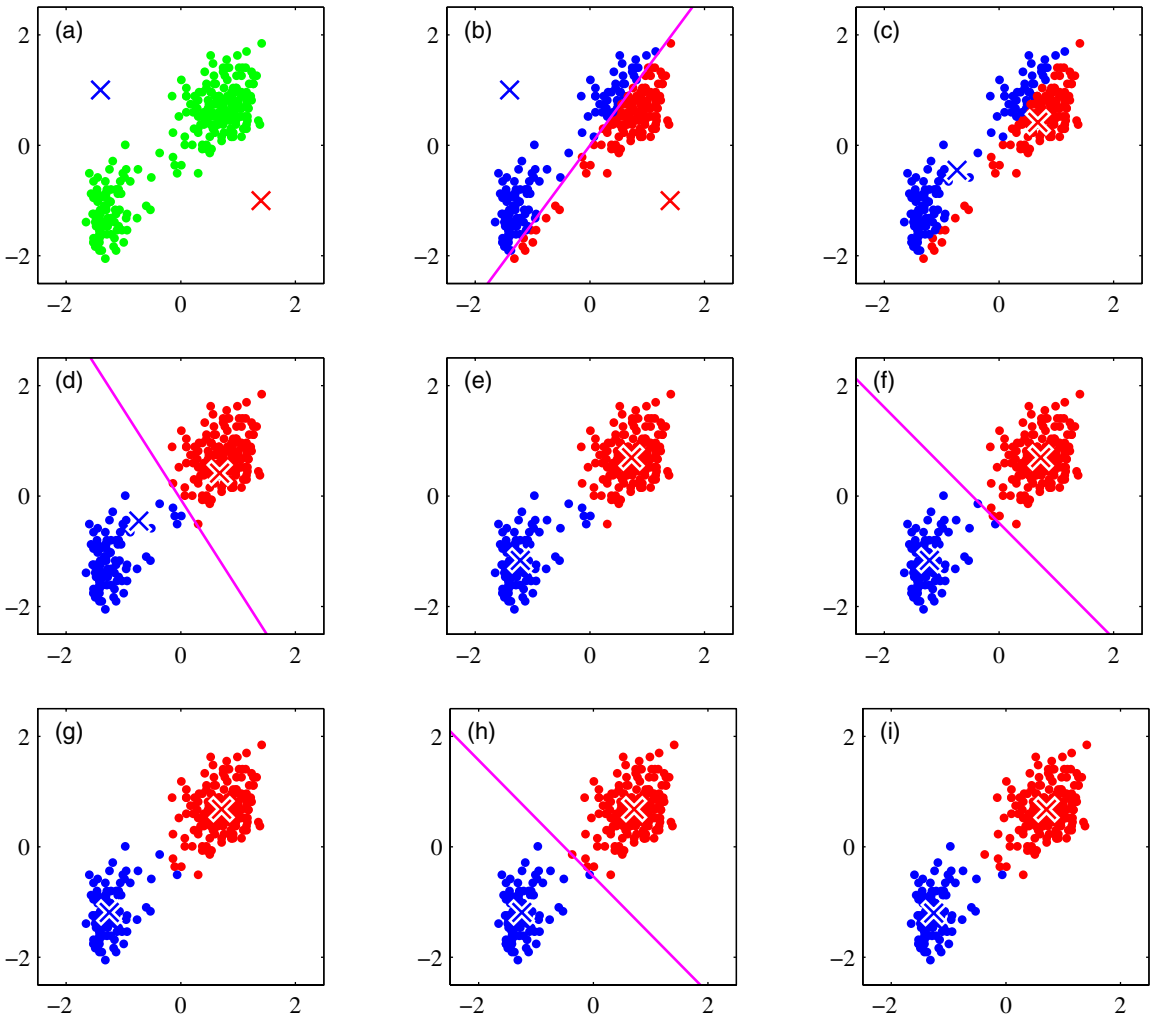
kmeans: Procedure

AICIP RESEARCH

- Step1: Begin with an arbitrary assignment of samples to clusters or begin with an arbitrary set of cluster centers and assign samples to nearest clusters
- Step2: Compute the sample mean of each cluster
- Step3: Reassign each sample to the cluster with the nearest mean
- Step4: If the classification of all samples has not changed, stop; else go to step 2.

kmeans: Illustration

AICIP RESEARCH



kmeans: Objective function

AICIP RESEARCH

- Find r_{nk} and $\boldsymbol{\mu}_k$ such that J is minimized
  - K: number of clusters
  - $\boldsymbol{\mu}_k$: cluster centers
  - N: number of data samples
  - r_{nk}: a mask

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2$$

- Optimization

$$r_{nk} = \begin{cases} 1 & \text{if } k = \arg\min_j \|\mathbf{x}_n - \boldsymbol{\mu}_j\|^2 \\ 0 & \text{otherwise.} \end{cases}$$

Exhaustive search

$$2 \sum_{n=1}^{N} r_{nk}(\mathbf{x}_n - \boldsymbol{\mu}_k) = 0 \qquad \boldsymbol{\mu}_k = \frac{\sum_n r_{nk} \mathbf{x}_n}{\sum_n r_{nk}}.$$

Newton's method

Winner-takes-all: Procedure

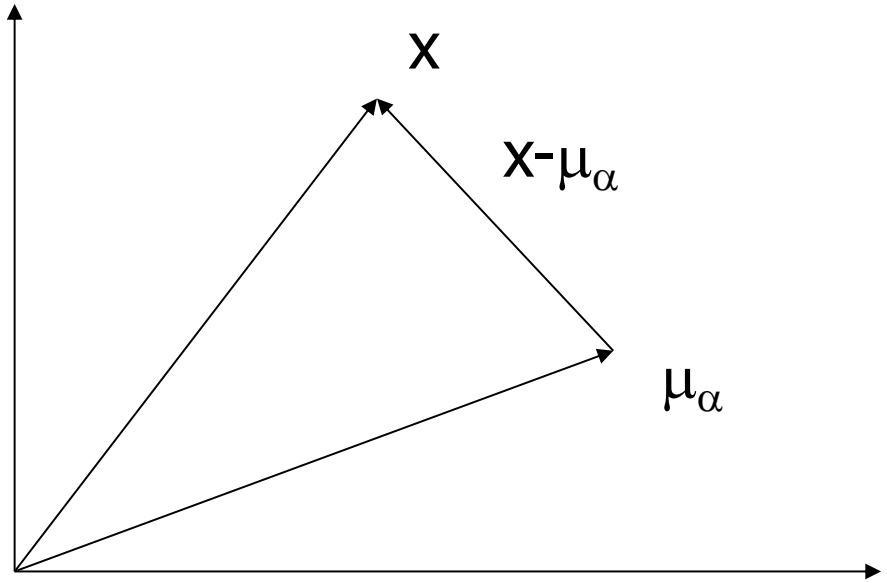
AICIP RESEARCH

- Step 1: Begin with an arbitrary set of cluster centers μ_i
- Step 2: For each sample $\boldsymbol{x}_n$, find the nearest cluster center μ_α, which is called the winner.
- Step 3: Modify μ_α using $\mu_\alpha^{new} = \mu_\alpha^{old} + \eta(\mathbf{x} - \mu_\alpha^{old})$
  - η is known as a "learning parameter".
  - Typical values of this parameter are small, on the order of 0.01.
- Step 4: If the classification of all samples has not changed, stop; else go to step 2.

Batch processing vs. On-line processing

Winner-takes-all: Illustration

AICIP RESEARCH



AICIP RESEARCH

Winner-takes-all: Objective function

- Same as kmeans

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2$$

- Optimization: Gradient Descent

$$\boldsymbol{\mu}_k^{\text{new}} = \boldsymbol{\mu}_k^{\text{old}} + \eta_n (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{old}})$$

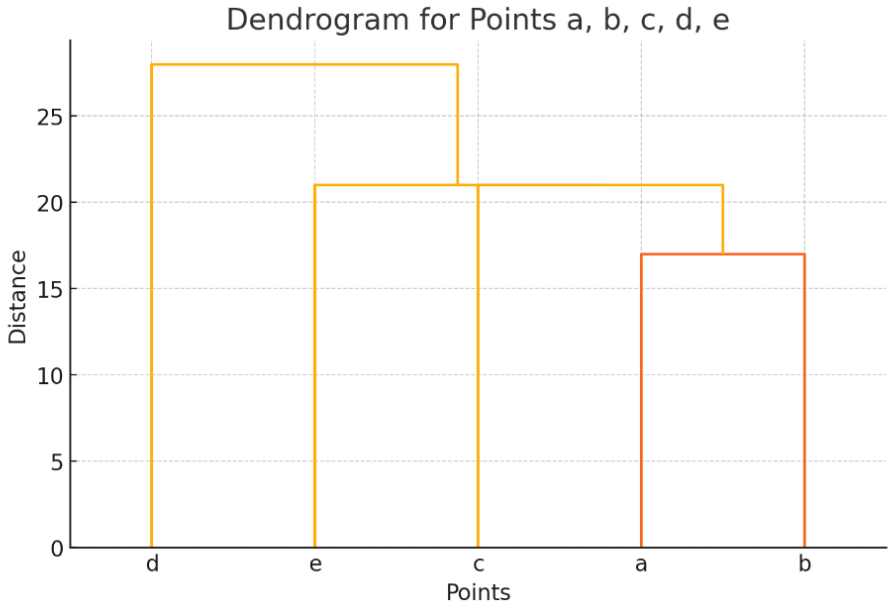
PART II: HIERARCHICAL CLUSTERING

Hierarchical Clustering Algorithm

AICIP RESEARCH

- Agglomerative clustering (bottom-up)
- Divisive (top-bottom)

- Dendrogram
- Cluster linkage
  - Single linkage: d_{min}
  - Complete linkage: d_{max}
  - Centroid linkage: d_{mean}
  - SLINK



AICIP RESEARCH

Agglomerative Hierarchical Clustering – Procedure

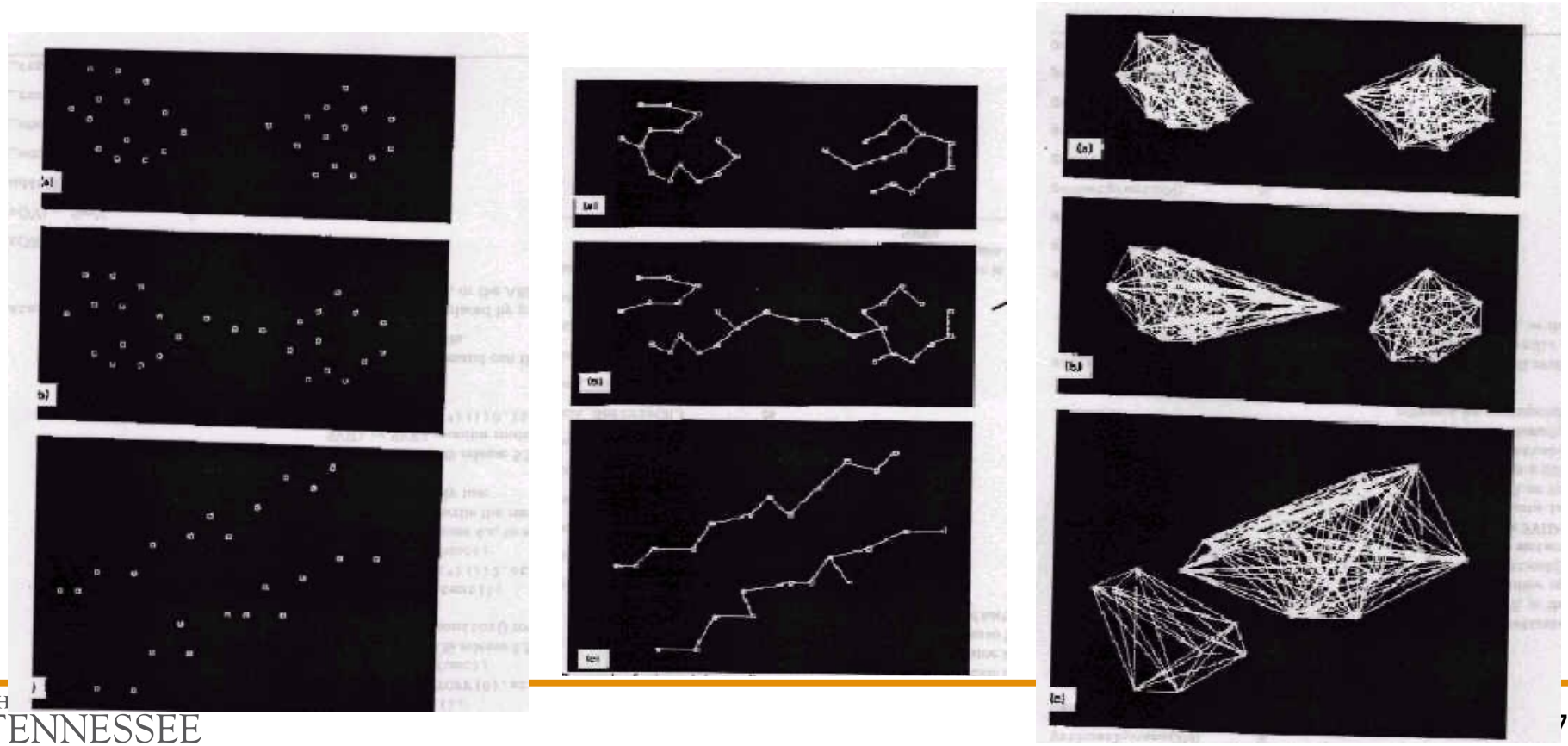
- Step1: assign each data point in the data set to a separate cluster
- Step2: merge the two "closest" clusters
- Step3: repeat step2 until you get the number of clusters you want or the appropriate cluster number

- The result is highly dependent on the measure of cluster distance (or the linkage function)

AICIP RESEARCH

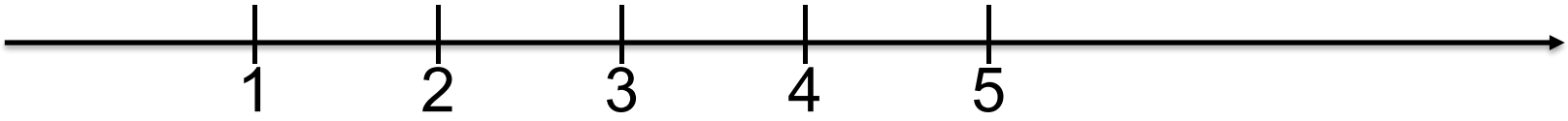
Comparison of Shape of Clusters

- dmin tends to choose clusters which are ??
- dmax tends to choose clusters which are ??



Agglomerative clustering: example 1

AICIP RESEARCH



Agglomerative clustering: example 2

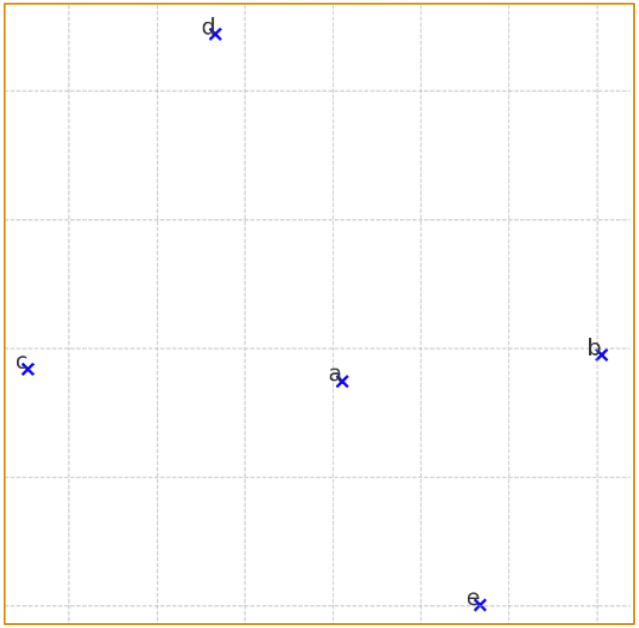
	a	b	c	d	e
a	0	17	21	31	23
b	17	0	30	34	21
c	21	30	0	28	39
d	31	34	28	0	43
e	23	21	39	43	0

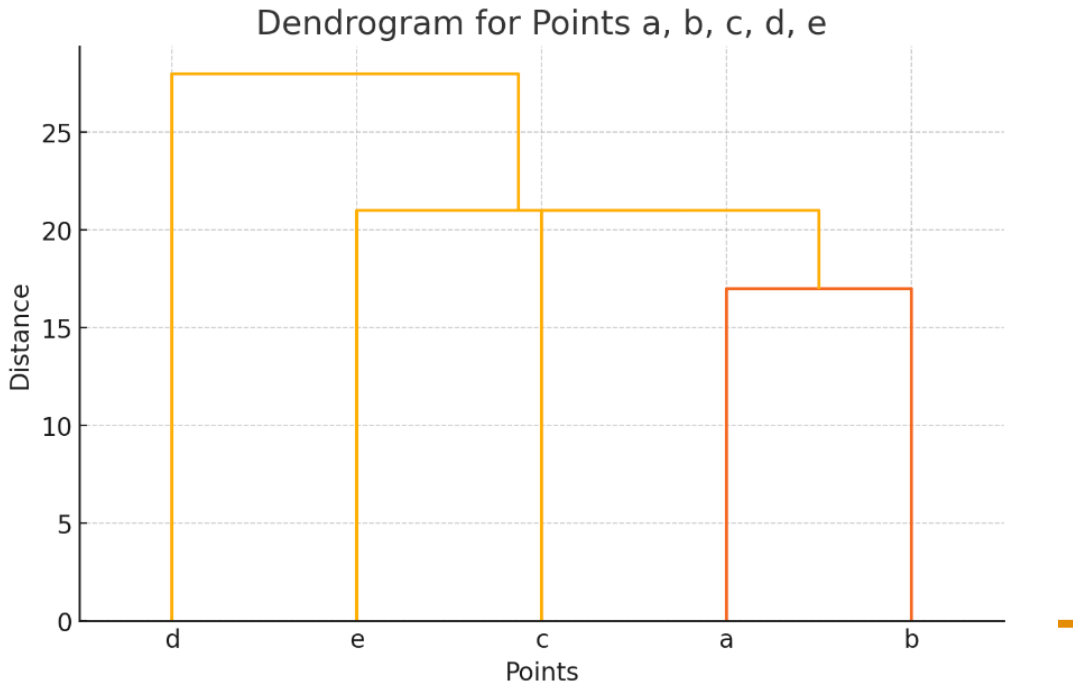
	(a,b)	c	d	e
(a,b)	0	**21**	**31**	**21**
c	**21**	0	*28*	*39*
d	**31**	*28*	0	*43*
e	**21**	*39*	*43*	0

	((a,b),c,e)	d
((a,b),c,e)	0	**28**
d	**28**	0

Distance Matrix

Dendrogram:
- Treelike
- Leaves
- Nodes
- Height





Example from https://en.wikipedia.org/wiki/Single-linkage_clustering

AICIP RESEARCH

Comparison: kmeans vs. hierarchical clustering

- Algorithm/procedure (cluster structure)
- # of clusters
- Cluster shape
- Deterministic?
- Complexity?
- Interpretability