CS5100: Data Analysis

Chapter 7: Unsupervised Learning - Clustering

Supervised Vs Unsupervised Learning

- In Supervised learning
  - all the datasets have labels
- The main target for data analysis on such datasets is
  - to fit a prediction function
  - which describes the relationship between labels and attributes:

$$y = \hat{f}(X)$$

- This is called supervised learning,
  - because it is a training process with a "teacher" telling you what is the correct "answer"

- But there are many datasets which do not have labels, for reasons:
  - different variables simply do not have a relationship
  - labelling large datasets can be costly

Unsupervised Learning

- The main target for data analysis on unlabelled datasets
  - is to understand the structure of the data:
  - Do the observations form clusters?
  - How are the observations generated in a random process?
  - Do the observations have hidden patterns, e.g. they are lying in a low-dimension manifold?

- This is called unsupervised learning.
  - Three main categories are - clustering
    - density estimation
    - and dimensionality reduction

Clustering

- Why clustering?

-some examples to show how clustering can be useful.

- Intuitively, clustering means classifying observations into different grops
- Sometimes, clusters are not that clear-cut, and two clusters may overlap.

Example ① Campaigning in US elections is expensive, so the parties are highly strategic in order to be efficient.

For instance, analysts will distinguish between

"old white males living in unwealthy states"

and

"highly educated voters living in swing states".

Note that: the two clusters do overlap.

Custering helps the campaigns to

- understand the demography of voters;
- Set up focus groups/ cold calls/ polls to know which political topics the different groups of voters concern.

Example ② Online shopping companies have a lot of data of
- customers' personal info
- browsing history and
- purchase history.

These companies are interested in recommending Marco with products he has never browsed before, so as to open new markets.

- In other words, if the "label" is a T/F variable
  - which indicates whether a customer is likely to buy a product,
  - the observation of Marco is unlabelled.

- The companies do clustering of all customers,
  - and see which clusters Marco belong to
  - (there may be more than one clusters).

- In each such cluster, the company finds out
  - the labelled observations (which are often not many),
  - and see if $T \gg F$ or not.

- if this happens, then the company is confident that
  - its recommendation will pursue Marco to buy that product.

K-Means Algorithm

- Distance between two observations
- Centroid of a set of observations
- The algorithm
- Discussions on the performance of the algorithm.

Data Model for Clustering

- The i^{th} observation is denoted by x_i,
  - which is a vector in $\mathbb{R}^d$, for some $d \geq 1$.

- For any two observation x_i, x_j
  - The Euclidean distance between them is

$$\|x_i - x_j\| = \sqrt{\sum_{k=1}^{d} (x_{ik} - x_{jk})^2}$$

Example ① $x_1 = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}$ and $x_2 = \begin{bmatrix} 7 \\ -3 \\ 6 \end{bmatrix}$, then

$$\|x_1 - x_2\| = \sqrt{(1-7)^2 + (-1+3)^2 + (3-6)^2} = \sqrt{49} = 7$$

- For any set of observations,
  - the centroid of them is simply their average.

Example ②

$$x_1 = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 7 \\ -3 \\ 6 \end{bmatrix}, \quad x_4 = \begin{bmatrix} -5 \\ 6 \\ 3 \end{bmatrix}, \quad x_8 = \begin{bmatrix} 3 \\ 6 \\ -2 \end{bmatrix} \in \mathbb{R}^3$$

Then the centroid of $\{x_1, x_2, x_4, x_8\}$ is

$$\text{Centroid} = \frac{1}{4} \begin{bmatrix} 1+7-5+3 \\ -1-3+6+6 \\ 3+6+3-2 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 2 \\ 2.5 \end{bmatrix}$$

K-Means Algorithm

- Input:
  - a dataset with observations in $\mathbb{R}^d$.
  - And an integer $K \geq 2$, K = max nº of clusters.

- Algorithm:

1. For each observation,
   - Assign it with a random integer $m \in [1, K]$,
   - which means the observation is initially assigned to cluster m

Then repeat the following two steps, until the clusters do not change anymore.

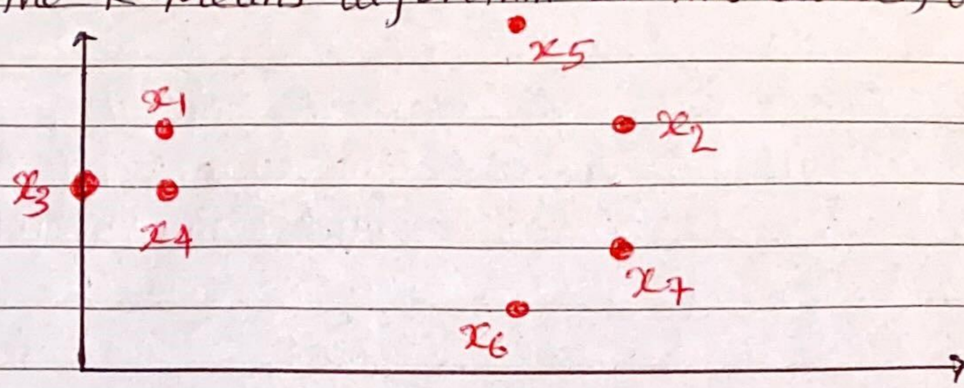
2. For each cluster, compute its centroid.
3. For each observation,
   - re-assign it to the cluster whose centroid is
   - the closest to the observation.

- If there is a tie, priority is given to staying with the current cluster.

Example 1 A dataset has 7 observations, each in $\mathbb{R}^2$:

$$\left\{ x_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, x_2 = \begin{bmatrix} 6 \\ 4 \end{bmatrix}, x_3 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, x_4 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, x_5 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}, x_6 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}, x_7 = \begin{bmatrix} 6 \\ 2 \end{bmatrix} \right\}$$

Run the K-Means algorithm on this dataset, with $K = 3$.



Step 1: Assign each observation to one of the 3 clusters randomly.

- Suppose the random assignment is given by the table below.

Observation	x_1	x_2	x_3	x_4	x_5	x_6	x_7
Cluster	3	2	2	1	3	1	2

Step 2: Compute the centroid of each cluster.

Centroid of cluster 1 $= \frac{1}{2}\begin{bmatrix} 1+5 \\ 3+1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 6 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

Centroid of cluster 2 $= \frac{1}{3}\begin{bmatrix} 6+0+6 \\ 4+3+2 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 12 \\ 9 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$

Centroid of cluster 3 $= \frac{1}{2}\begin{bmatrix} 1+5 \\ 4+6 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 6 \\ 10 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$

Step ③ For each observation,
- Re-assign it to the cluster whose centroid
- is the closet to the observation.

- If there is a tie, priority is given to staying with the current cluster.

- For convenience, we create the following table.
- In each box, we compute the square of the Euclidean distance between an observation & a centroid.

Iteration 1.

Clusters:	Cluster 1	Cluster 2	Cluster 3	Re-assign to cluster
Centroids:	$\begin{bmatrix}3\\2\end{bmatrix}$	$\begin{bmatrix}4\\3\end{bmatrix}$	$\begin{bmatrix}3\\5\end{bmatrix}$	
$x_1=\begin{bmatrix}1\\4\end{bmatrix}$	$(1-3)^2+(4-2)^2=8$	$(1-4)^2+(4-3)^2=10$	$(1-3)^2+(4-5)^2=5$	3
$x_2=\begin{bmatrix}6\\4\end{bmatrix}$	$(6-3)^2+(4-2)^2=13$	$(6-4)^2+(4-3)^2=5$	10	2
$x_3=\begin{bmatrix}0\\3\end{bmatrix}$	10	16	13	1
$x_4=\begin{bmatrix}1\\3\end{bmatrix}$	5	9	8	1
$x_5=\begin{bmatrix}5\\6\end{bmatrix}$	20	10	5	3
$x_6=\begin{bmatrix}5\\1\end{bmatrix}$	5	5	20	1
$x_7=\begin{bmatrix}6\\2\end{bmatrix}$	9	5	18	2

- The cluster assignment changes from

Observation	x_1	x_2	x_3	x_4	x_5	x_6	x_7
Cluster	3	2	2	1	3	1	2

to

x_1	x_2	x_3	x_4	x_5	x_6	x_7
3	2	1	1	3	1	2

As there is change, we need to repeat Steps 2 and 3.

$$\text{Centroids of cluster 1} = \frac{1}{3}\begin{bmatrix}0+1+5\\3+3+1\end{bmatrix} = \frac{1}{3}\begin{bmatrix}6\\7\end{bmatrix} = \begin{bmatrix}2\\7/3\end{bmatrix}$$

$$\text{Centroids of cluster 2} = \frac{1}{2}\begin{bmatrix}6+6\\4+2\end{bmatrix} = \frac{1}{2}\begin{bmatrix}12\\6\end{bmatrix} = \begin{bmatrix}6\\3\end{bmatrix}$$

$$\text{Centroids of cluster 3} = \frac{1}{2}\begin{bmatrix}1+5\\4+6\end{bmatrix} = \frac{1}{2}\begin{bmatrix}6\\10\end{bmatrix} = \begin{bmatrix}3\\5\end{bmatrix}$$

Iteration 2.

Clusters:	cluster 1	cluster 2	cluster 3	Re-assign to cluster
Centroids	$\begin{bmatrix} 2 \\ 7/3 \end{bmatrix}$	$\begin{bmatrix} 6 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 5 \end{bmatrix}$	
$x_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$	34/9	26	5	1
$x_2 = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$	25/9	1	10	2
$x_3 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$				1
$x_4 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$				1
$x_5 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$				3
$x_6 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$				2
$x_7 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$				2

- The cluster assignment changes from

observation	x_1	x_2	x_3	x_4	x_5	x_6	x_7
cluster	3	2	1	1	3	1	2

to

x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	2	1	1	3	2	2

- As there is change, we need to repeat steps 2 and 3.

Centroids of cluster 1: $\frac{1}{3}\begin{bmatrix} 1+0+1 \\ 4+3+3 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 \\ 10 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 10/3 \end{bmatrix}$

Centroids of cluster 2: $\frac{1}{3}\begin{bmatrix} 6+5+6 \\ 4+1+2 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 17 \\ 7 \end{bmatrix} = \begin{bmatrix} 17/3 \\ 7/3 \end{bmatrix}$

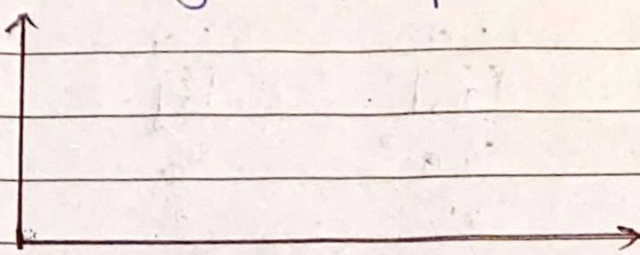
Centroids of cluster 3 : $\begin{bmatrix} 5 \\ 6 \end{bmatrix}$

- Interation 3

Clusters:	cluster 1	cluster 2	cluster 3	Re-assign to cluster
Centroids				
				1
				2
				1
				1
				3
				2
				2

- The cluster assignment does not change. So we can terminate, and output the final cluster assignment

- The final cluster assignment, represented on a plot:



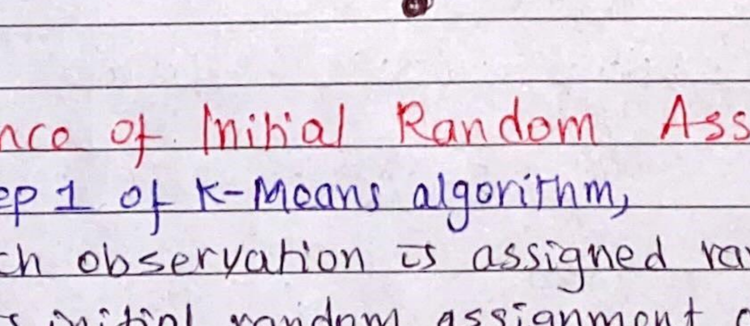
Discussions of K-Means Algorithm

Intuitions

- K-Means algorithm is an iterative algorithm.
  - In each iteration, it improves the total distance between each observation and the centroid of its cluster.

- Intuitively,
  - when each observation is close to the centroid of its cluster,
  - the observations in the same cluster are close to each other.

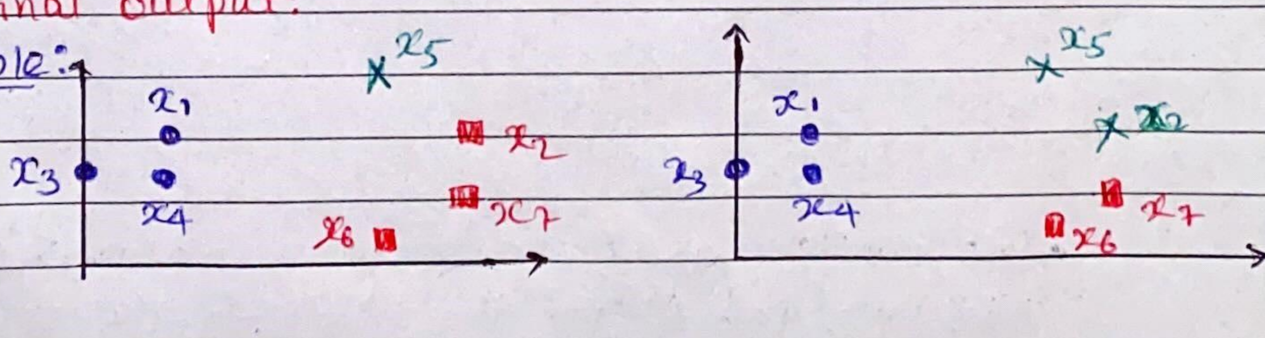
■: centroid of the cluster ●: observations in the cluster



Influence of Initial Random Assignment

- In Step 1 of K-Means algorithm,
  - each observation is assigned randomly to a cluster.
  - This initial random assignment does affect the final output.

Example:

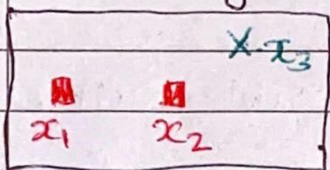


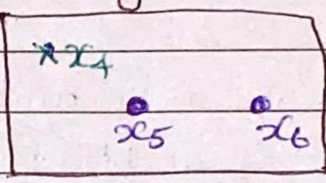
Less than K Clusters in the Output

- The parameter K specifies
  - the maximum possible number of clusters in the output.
- It is possible that less than K clusters are in the output.

Example: Suppose the initial assignment is given in figure below.

- After the 1st iteration of K-Means,
  - x_3 will join the red cluster,
  - while x_4 joins the blue cluster.
  - The green cluster no longer exists.





K-Means Must Terminate

- When we design an algorithm,
  - one key aspect is its running time, or
  - whether it will terminate on any legitimate input.
- It is not clear that K-Means must terminate.

Can several cluster assignments keep looping in a cycle during the execution of K-Means?

- We will show that each cluster assignment appears at most once (except in the last iteration).
  - we use the following potential function.
  - For each cluster assignment C,
  - let C_j be its j^{th} cluster, and let V_j denote the centroid of C_j.

$$\Phi(C) = \sum_{C_j} \sum_{x \in C_j} \|x - V_j\|^2$$

In other words, $\Phi(C)$ is the total distance between each observation and its cluster's centroid.

Theorem

(i) In K-Means,

- the value of $\Phi(C)$ strictly decreases after each iteration,
- except for the last iteration.

ii) Consequently, K-Means must terminate on any input.

Proof of (ii)

- Part (i) implies that no cluster assignment appears twice
  - during an execution of K-Means
  - (except the last iteration)

Since there are only finitely many possible cluster assignments, so K-Means must terminate.

- possible cluster assignments $= K^n$

Proof of (i)

Recall that K-Means runs the two steps below in each iteration:

2. For each cluster, compute its centroid.
3. For each observation,
   - Re-assign it to the cluster
   - whose centroid is closest to the observation.

(Priority is given to staying with the current cluster.)

In step 2, the clusters are not changed.

- The centroid of a cluster C_j is the point v that minimizes

$$\sum_{x \in C_j} \|x - v\|^2$$

- Thus, for every j, $\sum_{x \in C_j} \|x - v_j\|^2$
  - does not increase after Step 2,
  - and hence $\Phi(C) = \sum_j \sum_{x \in C_j} \|x - v_j\|^2$ does not increase.

Recall that: $\Phi(C)$ is the total distance between

- each observation and its cluster's centroid.

In step 3, whenever an observation is re-assigned,

- its distance from its cluster's centroid strictly decreases
- (you need the tie-breaking rule to ensure this).

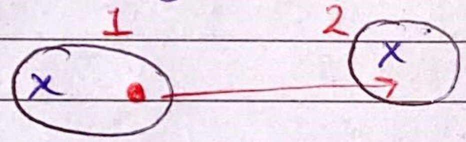
Thus, $\Phi(C)$ strictly decreases whenever an observation being

- re-assigned

When no observation is re-assigned in an iteration,

- $\Phi(C)$ stays the same.

- Re-assignment of an observation



- reduces its distance from its cluster's centroid.

Challenge!!!

Retake

Week 7:

Clustering

Hierarchical Clustering (HC)

- Distance between two clusters
  - single linkage
  - complete linkage
  - average linkage and
  - centroid linkage
- Agglomerative Hierarchical clustering Algorithm
- Dendrogram

Hierarchical Clustering (HC)

- In K-Means, the parameter K is specified of the iterations, and at most K clusters are produced.
  - Also, the output depends on the random initial assignment.
  - It will be better if there is a method whose output is deterministic, and we can choose
- Hierarchical clustering is one such method,
  - by arranging clusters & sub-clusters in tree structure.
- To perform HC, we need to define a distance between clusters.

"There" are 4 choices.

1. Single linkage
   - Given two clusters denoted by ω_1, ω_2, the single linkage between the two clusters is

$$d_{min}(\omega_1, \omega_2) = \min_{\substack{x_1 \in \omega_1 \\ x_2 \in \omega_2}} \|x_1 - x_2\|$$

2. Complete linkage

- The complete linkage between the two clusters is

$$d_{max}(\omega_1, \omega_2) = \max_{\substack{x_1 \in \omega_1 \\ x_2 \in \omega_2}} \|x_1 - x_2\|$$

3. Average linkage

- The average linkage between the two clusters is

$$d_{avg}(\omega_1, \omega_2) = \frac{1}{|\omega_1| \cdot |\omega_2|} \sum_{x \in \omega_1} \sum_{x_2 \in \omega_2} \|x_1 - x_2\|$$

4. Centroid linkage

- The centroid linkage between the two clusters is the distance between their centroids.

Example: Suppose $\omega_1 = \left\{ \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \end{bmatrix} \right\}$ & $\omega_2 = \left\{ \begin{bmatrix} 5 \\ 2 \end{bmatrix}, \begin{bmatrix} 6 \\ 4 \end{bmatrix}, \begin{bmatrix} 7 \\ 3 \end{bmatrix} \right\}$

What are the single, complete, average & centroid linkage between the two clusters?

1. single linkage $d_{min}(\omega_1, \omega_2) = \min_{\substack{x_1 \in \omega_1 \\ x_2 \in \omega_2}} \|x_1 - x_2\|$

- The Euclidean distance

$$\|x_1 - x_2\| = \sqrt{(1-5)^2 + (4-2)^2} = \sqrt{(-4)^2 + (2)^2} = \sqrt{16+4} = \sqrt{20} = 2\sqrt{5}$$

$\|x_1 - x_2\| =$

	$\begin{bmatrix} 5 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 6 \\ 4 \end{bmatrix}$	$\begin{bmatrix} 7 \\ 3 \end{bmatrix}$
$\begin{bmatrix} 1 \\ 4 \end{bmatrix}$	$\sqrt{20}$	5	$\sqrt{37}$
$\begin{bmatrix} 3 \\ 2 \end{bmatrix}$	2	$\sqrt{13}$	$\sqrt{17}$

Single linkage $d_{min}(\omega_1, \omega_2) = 2$

Complete linkage $d_{max}(\omega_1, \omega_2) = \sqrt{37} = 6.0828$

Average linkage $d_{avg}(\omega_1, \omega_2) = \frac{1}{6}\left[\sqrt{20} + 5 + \sqrt{37} + 2 + \sqrt{13} + \sqrt{17}\right)$

$= 4.2139$

Agglomerative HC Algorithm

- Input: A dataset with n observations on $\mathbb{R}^d$.
  Choose a linkage.
- Algorithm:
  1. Each observation forms a cluster size 1.
     Thus, there are n clusters at the beginning.

  Then repeat the following two steps, until there is only one cluster left.

  2. For any pair of remaining clusters, compute the distance between them using the linkage you chose.

  3. Merge the two clusters that achieve the minimum distance into one cluster.
     → The distance between them will become the height of their parent in the dendrogram.

Example: A dataset has seven observations, each in $\mathbb{R}^2$:

$$\left\{x_1 = \begin{bmatrix}1\\4\end{bmatrix}, x_2 = \begin{bmatrix}6\\4\end{bmatrix}, x_3 = \begin{bmatrix}0\\3\end{bmatrix}, x_4 = \begin{bmatrix}1\\3\end{bmatrix}, x_5 = \begin{bmatrix}5\\6\end{bmatrix}, x_6 = \begin{bmatrix}5\\1\end{bmatrix}, x_7 = \begin{bmatrix}6\\2\end{bmatrix}\right\}$$

We will run the agglomerative HC algorithm on this data set, using the complete linkage.

First, each observation forms a cluster by itself:

$$\omega_1 = \{x_1\}, \omega_2 = \{x_2\}, \ldots, \omega_7 = \{x_7\}$$

Second, for each pair of clusters
- we compute the complete linkage between them.

Mar.

$x_1 = \begin{bmatrix}1\\4\end{bmatrix}$ $x_2 = \begin{bmatrix}6\\4\end{bmatrix}$ $x_4 = \begin{bmatrix}1\\3\end{bmatrix}$

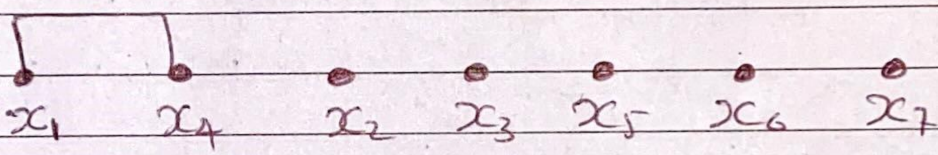
Iteration 1:

	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_4\}$	$\{x_5\}$	$\{x_6\}$	$\{x_7\}$
$\{x_1\}$	-	$\sqrt{25}$	$\sqrt{2}$	$\sqrt{1}$	$\sqrt{20}$	$\sqrt{25}$	$\sqrt{29}$
$\{x_2\}$	-	-	$\sqrt{37}$	$\sqrt{26}$	$\sqrt{5}$	$\sqrt{10}$	$\sqrt{4}$
$\{x_3\}$	-	-	-	$\sqrt{1}$	$\sqrt{34}$	$\sqrt{29}$	$\sqrt{37}$
$\{x_4\}$	-	-	-	-	$\sqrt{25}$	$\sqrt{20}$	$\sqrt{26}$
$\{x_5\}$	-	-	-	-	-	$\sqrt{25}$	$\sqrt{17}$
$\{x_6\}$	-	-	-	-	-	-	$\sqrt{2}$
$\{x_7\}$	-	-	-	-	-	-	-

$$\|x_1 - x_2\| = \sqrt{(1-6)^2 + (4-4)^2} = \sqrt{(-5)^2+0} = \sqrt{25}$$

$$d_{max}(\omega_1, \omega_2) = \max_{\substack{x_1 \in \omega_1 \\ x_2 \in \omega_2}} \|x_1 - x_2\|$$

height $= \sqrt{1} = 1$



Iteration 2:

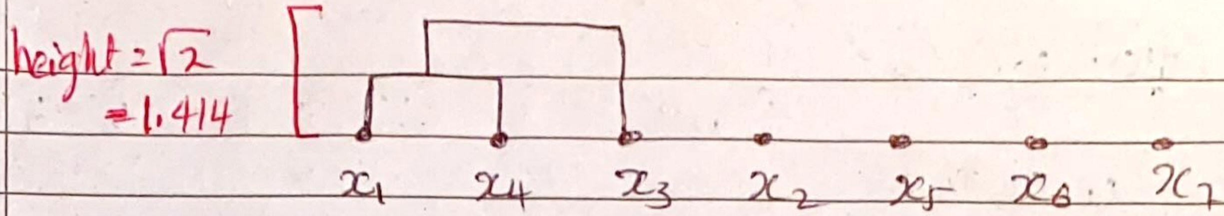
	$\{x_1, x_4\}$	$\{x_2\}$	$\{x_3\}$	$\{x_5\}$	$\{x_6\}$	$\{x_7\}$
$\{x_1, x_4\}$	-	$\sqrt{26}$	$\sqrt{2}$	$\sqrt{25}$	$\sqrt{25}$	$\sqrt{29}$
$\{x_2\}$	-	-	$\sqrt{37}$	$\sqrt{5}$	$\sqrt{10}$	$\sqrt{4}$
$\{x_3\}$	-		-	$\sqrt{34}$	$\sqrt{29}$	$\sqrt{37}$
$\{x_5\}$	-	-	-	-	$\sqrt{25}$	$\sqrt{17}$
$\{x_6\}$	-	-	-	-	-	$\sqrt{2}$
$\{x_7\}$	-	-	-	-	-	-

$$\|\{x_1,x_4\} - x_2\| = \sqrt{(2-6)^2 + (7-4)^2} = \sqrt{(-4)^2+3^2} = \sqrt{16+9} = \sqrt{25}$$

$$(x_1, x_4) = x_1 + x_4 = \begin{bmatrix}1\\4\end{bmatrix} + \begin{bmatrix}1\\3\end{bmatrix} = \begin{bmatrix}2\\7\end{bmatrix}$$

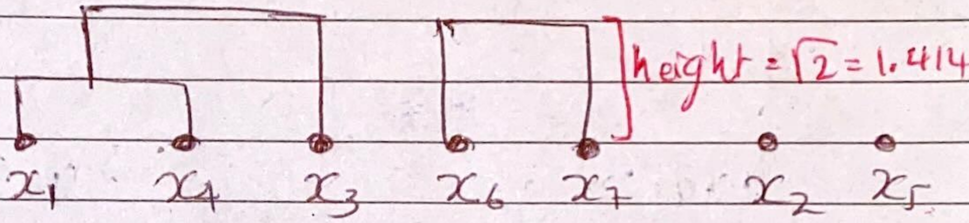
$\|x_2 - x_4\| = \qquad = \sqrt{26}$

Nota that:



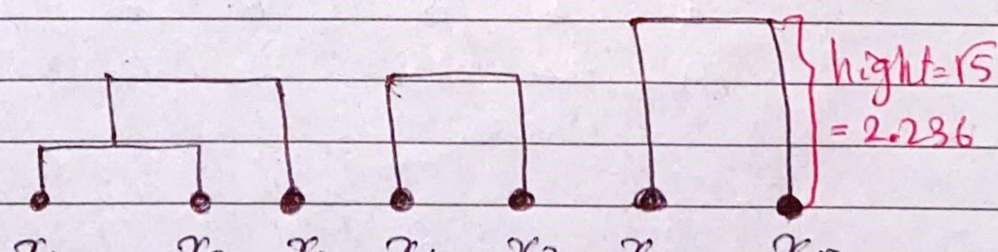
Iteration 3:

	$\{x_1, x_3, x_4\}$	$\{x_2\}$	$\{x_5\}$	$\{x_6\}$	$\{x_7\}$
$\{x_1, x_3, x_4\}$	–	$\sqrt{37}$	$\sqrt{34}$	$\sqrt{29}$	$\sqrt{37}$
$\{x_2\}$	–	–	$\sqrt{5}$	$\sqrt{10}$	$\sqrt{4}$
$\{x_5\}$	–	–	–	$\sqrt{25}$	$\sqrt{17}$
$\{x_6\}$	–	–	–	–	$\sqrt{2}$
$\{x_7\}$	–	–	–	–	–



Iteration 4:

	$\{x_1, x_3, x_4\}$	$\{x_2\}$	$\{x_5\}$	$\{x_6, x_7\}$
$\{x_1, x_3, x_4\}$	–	$\sqrt{37}$	$\sqrt{34}$	$\sqrt{37}$
$\{x_2\}$	–	–	$\sqrt{5}$	$\sqrt{10}$
$\{x_5\}$	–	–	–	$\sqrt{25}$
$\{x_6, x_7\}$	–	–	–	–



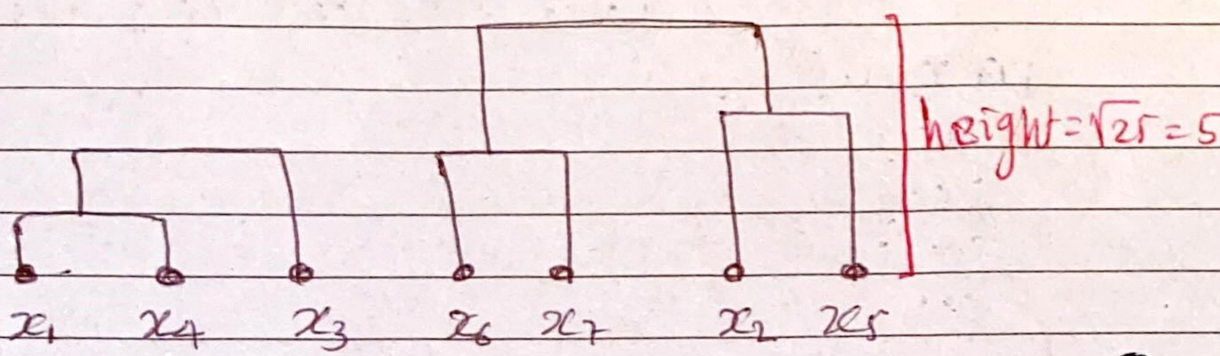
Iteration 5:

$\{x_1, x_3, x_4\}$ $\{x_2, x_5\}$ $\{x_6, x_7\}$

$\{x_1, x_3, x_4\}$

$\{x_2, x_5\}$ $\sqrt{37}$

$\{x_6, x_7\}$ $\sqrt{37}$ $\sqrt{25}$

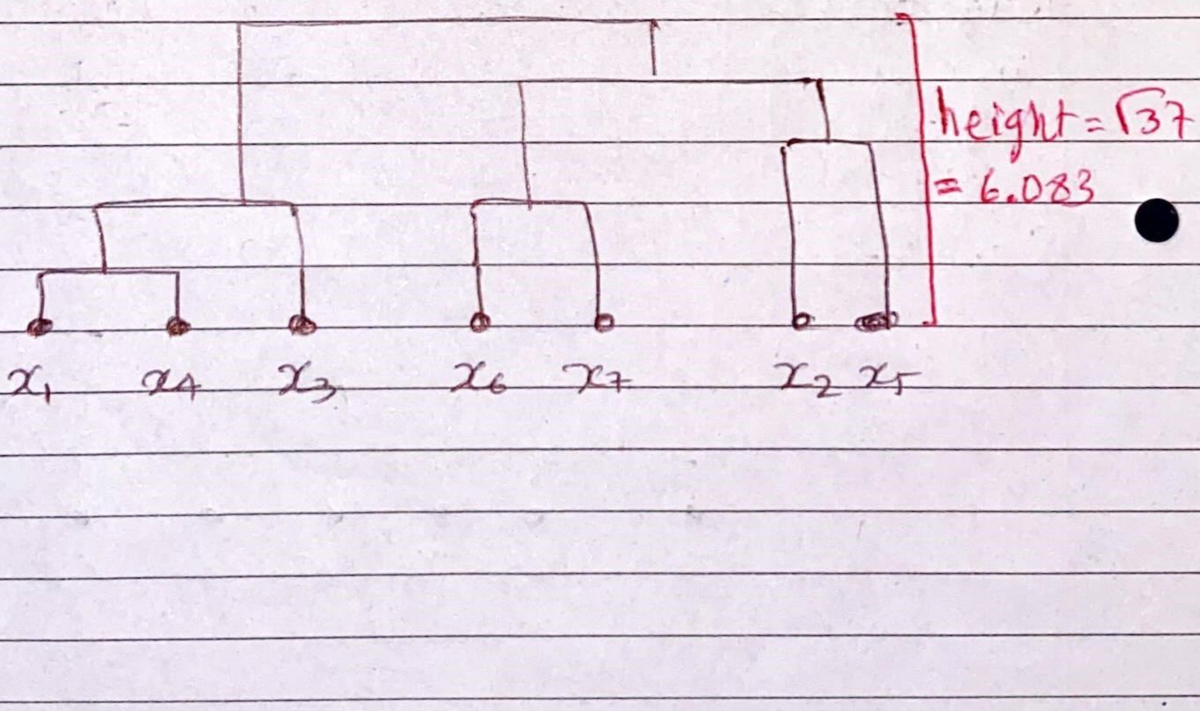


Iteration 6:

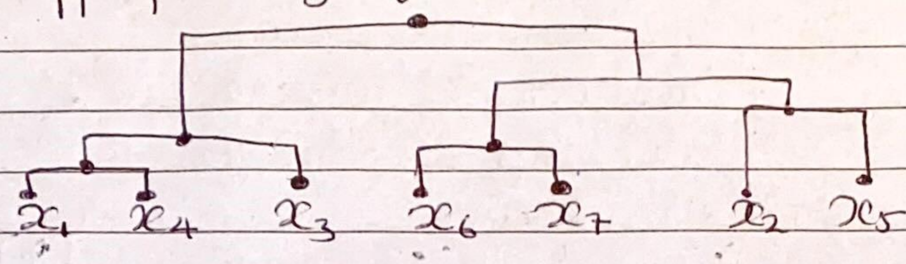
$\{x_1, x_3, x_4\}$ $\{x_2, x_5, x_6, x_7\}$

$\{x_1, x_3, x_4\}$ – –

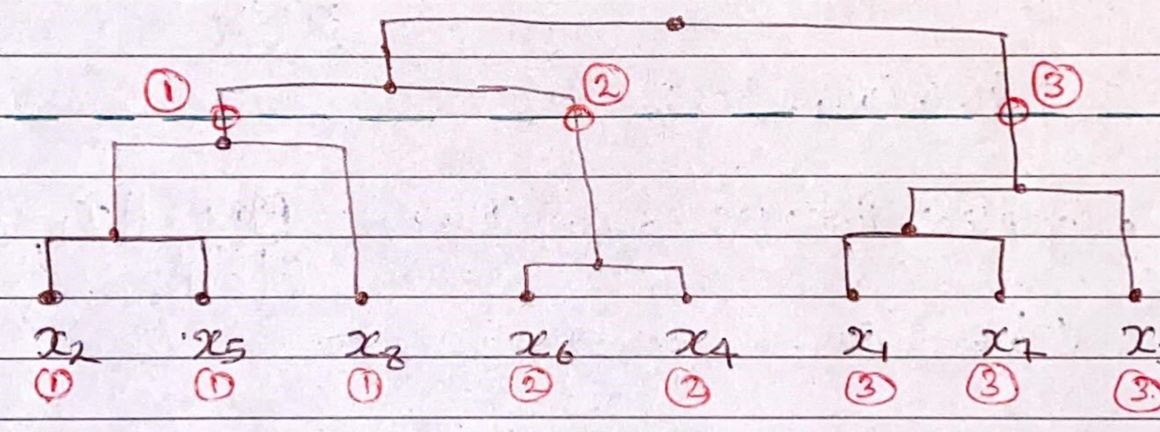
$\{x_2, x_5\}$ $\sqrt{37}$ –



Quiz 3: If we want 4 clusters by cutting the tree appropriately, what are the clusters?



Example: Dendrogram representing clusters in a tree structure.



Q: What should we do if we want to get 3 clusters out of the dendrogram?

A: Cut the tree at an appropriate height such that 3 clusters are formed.

The answer is $\{x_2, x_5, x_8\}$, $\{x_4, x_6\}$ and $\{x_1, x_3, x_7\}$.

Example: Suppose $\omega_1 = \left\{ \begin{bmatrix}1\\4\end{bmatrix}, \begin{bmatrix}3\\2\end{bmatrix} \right\}$ and $\omega_2 = \left\{ \begin{bmatrix}5\\2\end{bmatrix}, \begin{bmatrix}6\\4\end{bmatrix}, \begin{bmatrix}7\\3\end{bmatrix} \right\}$.

what are
- the single linkage,
- the complete linkage,
- the average linkage and
- the centroid linkage

between the two clusters?

	$\begin{bmatrix}5\\2\end{bmatrix}$	$\begin{bmatrix}6\\4\end{bmatrix}$	$\begin{bmatrix}7\\3\end{bmatrix}$
$\begin{bmatrix}1\\4\end{bmatrix}$	$\sqrt{(5-1)^2+(2-4)^2} = \sqrt{20}$	$\sqrt{5^2+0^2} = 5$	$\sqrt{6^2+1^2} = \sqrt{37}$
$\begin{bmatrix}3\\2\end{bmatrix}$	$\sqrt{(5-3)^2+(2-2)} = 2$	$\sqrt{3^2+2^2} = \sqrt{13}$	$\sqrt{4^2+1^2} = \sqrt{17}$

- Single linkage: $d_{min}(\omega_1, \omega_2) = \min\limits_{\substack{x_1 \in \omega_1 \\ x_2 \in \omega_2}} \|x_1 - x_2\|$

  $\therefore d_{min}(\omega_1, \omega_2) = 2.$

- Complete linkage: $d_{max}(\omega_1, \omega_2) = \max\limits_{\substack{x_1 \in \omega_1 \\ x_2 \in \omega_2}} \|x_1 - x_2\|$

  $\therefore d_{max}(\omega_1, \omega_2) = \sqrt{37} \approx 6.0828$

- Average linkage: $\dfrac{1}{|\omega_1| \cdot |\omega_2|} \sum\limits_{x \in \omega_1} \sum\limits_{x_2 \omega_2} \|x_1 - x_2\|$

  $d_{avg} = (\omega_1, \omega_2) = \dfrac{1}{2 \times 3} \left[\sqrt{20} + 5 + \sqrt{37} + 2 + \sqrt{13} + \sqrt{17} \right)$

  $= 4.21394.$

- The centroid linkage between the two clusters
  - is the distance between their centroids.

  centroid of $\omega_1 = \frac{1}{2}\begin{bmatrix}1+3\\4+2\end{bmatrix} = \frac{1}{2}\begin{bmatrix}4\\6\end{bmatrix} = \begin{bmatrix}2\\3\end{bmatrix}$

  centroid of $\omega_2 = \frac{1}{3}\begin{bmatrix}5+6+7\\2+4+3\end{bmatrix} = \frac{1}{3}\begin{bmatrix}18\\9\end{bmatrix} = \begin{bmatrix}6\\3\end{bmatrix}$

  centroid linkage b/n ω_1 & $\omega_2 = \sqrt{(2-6)^2 + (3-3)^2} = \sqrt{4^2} = 4$

Quiz 2: Suppose $\omega_1 = \left\{ \begin{bmatrix} 12 \\ 7 \end{bmatrix}, \begin{bmatrix} 11 \\ 5 \end{bmatrix} \right\}$ and $\omega_2 = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 3 \end{bmatrix} \right\}$.

What is the complete linkage between the two clusters?

	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -2 \\ 3 \end{bmatrix}$
$\begin{bmatrix} 12 \\ 7 \end{bmatrix}$	$\sqrt{(12-0)^2+(7-0)^2} = \sqrt{193}$	$\sqrt{14^2+4^2} = \sqrt{212}$
$\begin{bmatrix} 11 \\ 5 \end{bmatrix}$	$\sqrt{11^2+5^2} = \sqrt{146}$	$\sqrt{13^2+8^2} = \sqrt{173}$

$d_{max}(\omega_1, \omega_2) = \sqrt{212} = 14.560$