

# Ch 12.1, 12.4: Unsupervised Learning & Clustering

## Lecture 32 - CMSE 381

Prof. Guanqun Cao

Michigan State University

::

Dept of Computational Mathematics, Science & Engineering

Wed, Nov 19, 2025

# Announcements

## Last time:

- Convolutional Neural Nets

## This lecture:

- Clustering (Just hierarchical clustering)

## Announcements:

- No more homework!
- 11/26: Project office hours, zoom only (link on [website](#)), send a message on slack!
- Fri 11/21: Review - submit your questions
- Monday 11/24: Exam 3
  - ▶ Content since 2nd Exam (Ch 7 and on)
  - ▶ One page (8.5x11) handwritten cheat sheet
  - ▶ no-internet Calculator

21	F	10/24	Polynomial & Step Functions	7.1-7.2	HW #5 Due Sun 10/26
22	M	10/27	Step Functions; Basis functions; Start Splines	7.2-7.4	
23	W	10/29	Regression Splines	7.4	HW #6 Due Sun 11/2
24	F	10/31	Decision Trees	8.1	
25	M	11/3	Random Forests	8.2.1, 8.2.2	HW #7 Due Sun 11/9
26	W	11/5	Maximal Margin Classifier	9.1	
27	F	11/7	SVC	9.2	HW #8 Due Sun 11/16
28	M	11/10	SVM	9.3, 9.4	
29	W	11/12	Single Layer NN	10.1	HW #9 Due Sun 11/23
30	F	11/13	Multi Layer NN	10.2	
31	M	11/17	CNN	10.3	Project Due
32	W	11/19	Unsupervised learning / clustering	12.1, 12.4	
33	F	11/21	<b>Review</b>		Honors Project Due
	M	11/24	<b>Midterm #3</b>		
	W	11/26	Virtual: Project Office Hours		
	F	11/28	Thanksgiving		
	M	12/1	Virtual: Project Office Hours		
	W	12/3	Virtual: Project Office Hours		
	F	12/5			
	M	12/8			
	W	12/10			
	F	12/12	<b>No final exam</b>		

# Section 1

## Unsupervised learning

# Supervised vs Unsupervised Learning

**Supervised**

**Unsupervised**

# Some examples of unsupervised problems

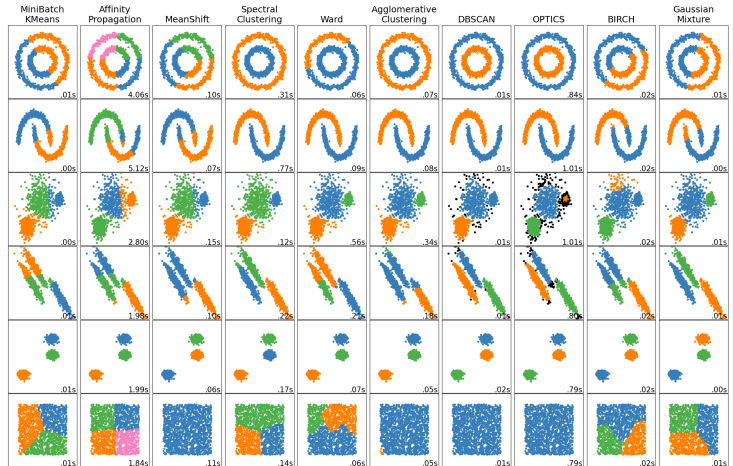
- Assay gene expression levels in 100 patients with breast cancer, looking for subgroups with similar qualities
- Online shopping: find groups of shoppers with similar browsing and purchase histories and show relevant related products.
- Search engine picking results to show

## Section 2

### Clustering

# Big idea

Clustering: relation between samples

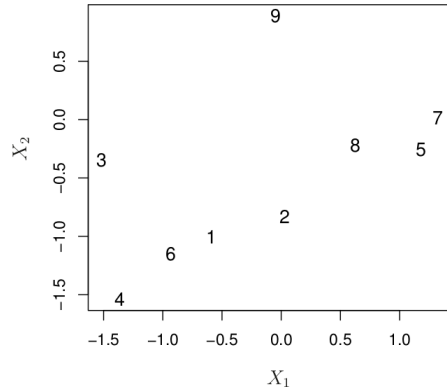
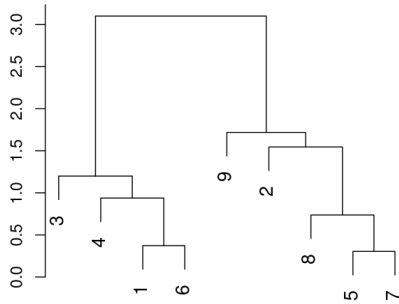


## Section 3

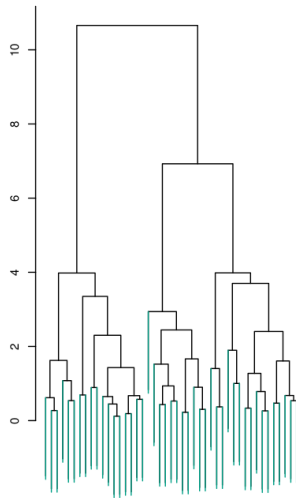
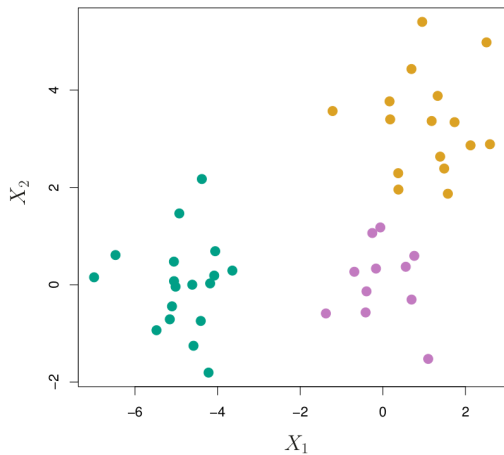
### Hierarchical Clustering



# Dendrogram



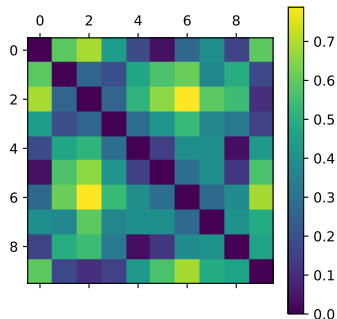
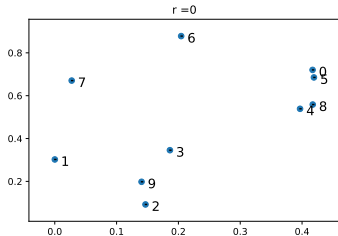
# A bigger example



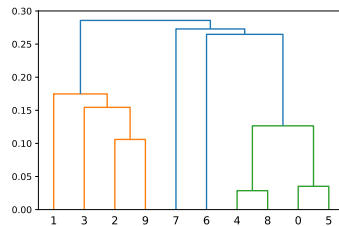
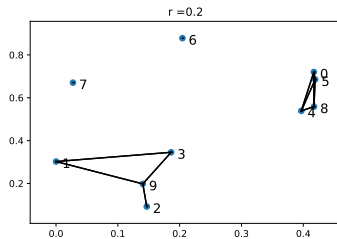
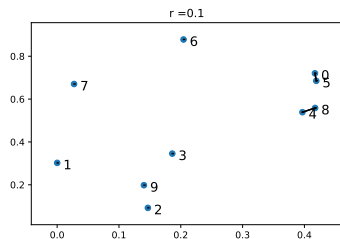
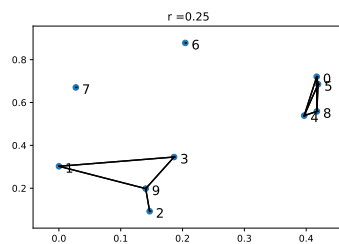
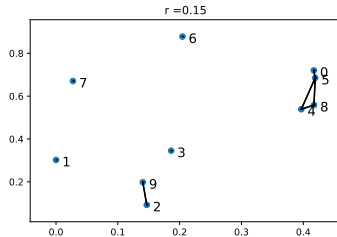
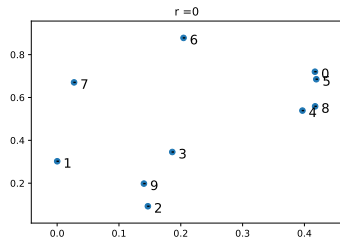
# Single linkage

Distance between cluster  $A$  and cluster  $B$ :  
Smallest distance between the points

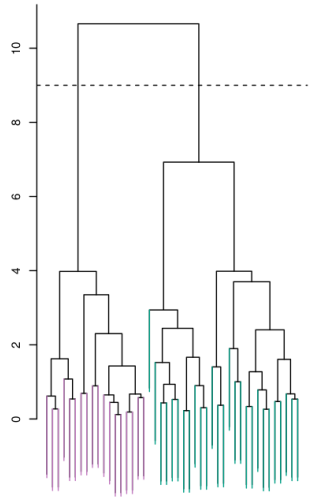
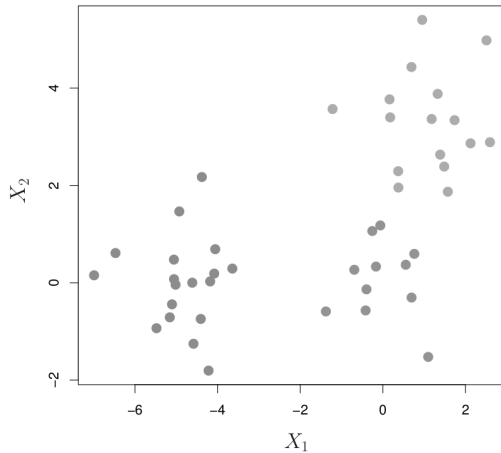
$$L(A, B) = \min_{a \in A, b \in B} \|a - b\|$$



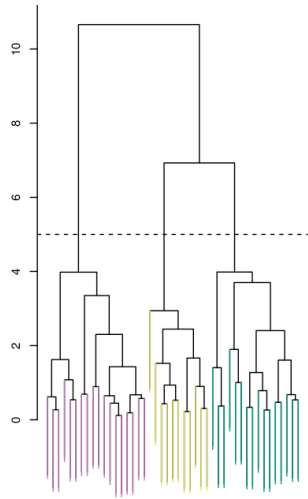
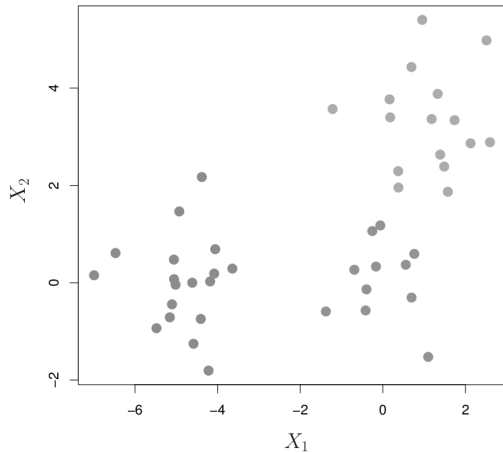
# Building the dendrogram



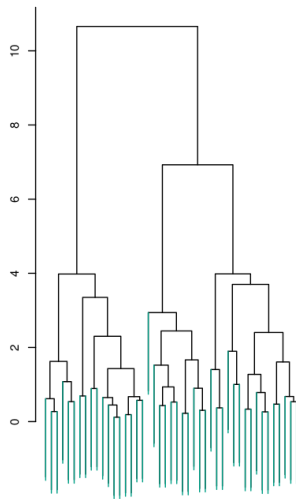
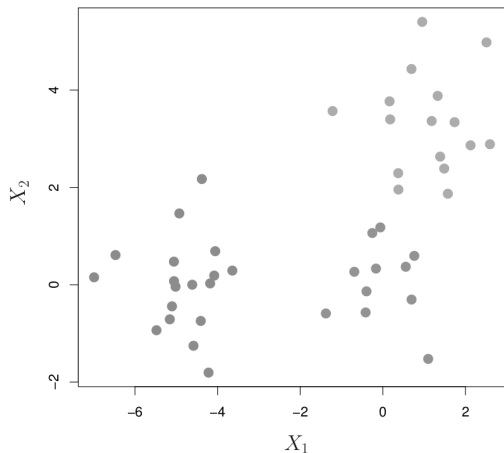
# How to get clusters



# How to get different clusters



# Can get any number of clusters



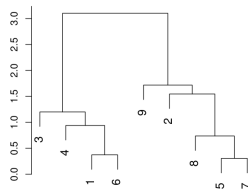
Test your understanding: [PolIEv](#)

# Linkage

<i>Linkage</i>	<i>Description</i>
Complete	Maximal intercluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B, and record the <i>largest</i> of these dissimilarities.
Single	Minimal intercluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B, and record the <i>smallest</i> of these dissimilarities. Single linkage can result in extended, trailing clusters in which single observations are fused one-at-a-time.
Average	Mean intercluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B, and record the <i>average</i> of these dissimilarities.
Centroid	Dissimilarity between the centroid for cluster A (a mean vector of length $p$ ) and the centroid for cluster B. Centroid linkage can result in undesirable <i>inversions</i> .

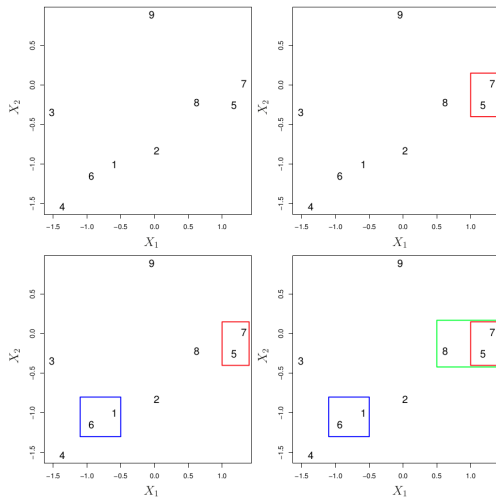


# Example with complete linkage

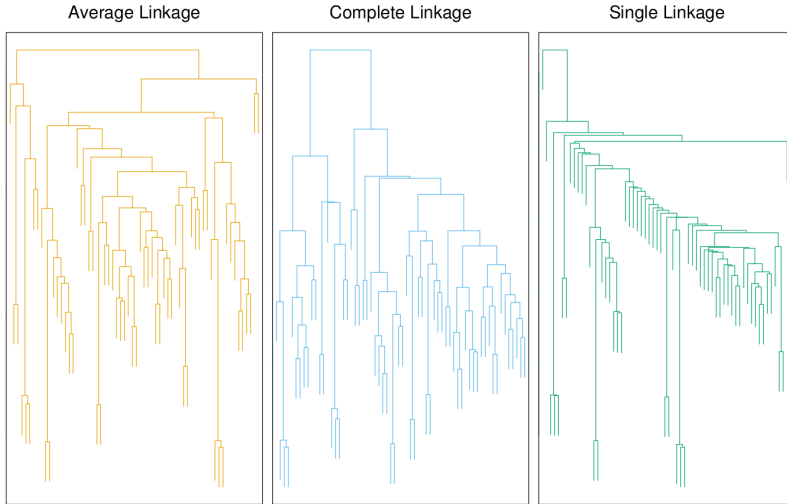


Distance between cluster  $A$  and cluster  $B$ :  
**Largest** distance between the points

$$L(A, B) = \max_{a \in A, b \in B} \|a - b\|$$



# Examples of different linkage



# Dependence on dissimilarity measure

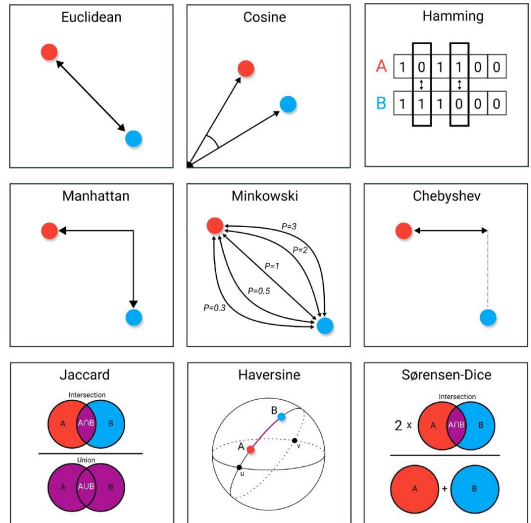


Photo Credit Link



# Next time

21	F	10/24	Polynomial & Step Functions	7.1-7.2	HW #5 Due Sun 10/26
22	M	10/27	Step Functions; Basis functions; Start Splines	7.2-7.4	
23	W	10/29	Regression Splines	7.4	
24	F	10/31	Decision Trees	8.1	HW #6 Due Sun 11/2
25	M	11/3	Random Forests	8.2.1, 8.2.2	
26	W	11/5	Maximal Margin Classifier	9.1	
27	F	11/7	SVC	9.2	HW #7 Due Sun 11/9
28	M	11/10	SVM	9.3, 9.4	
29	W	11/12	Single Layer NN	10.1	
30	F	11/13	Multi Layer NN	10.2	HW #8 Due Sun 11/16
31	M	11/17	CNN	10.3	
32	W	11/19	Unsupervised learning / clustering	12.1, 12.4	
33	F	11/21	<b>Review</b>		HW #9 Due Sun 11/23
	M	11/24	<b>Midterm #3</b>		
	W	11/26	Virtual: Project Office Hours		
	F	11/28	Thanksgiving		
	M	12/1	Virtual: Project Office Hours		
	W	12/3	Virtual: Project Office Hours		
	F	12/5			<b>Project Due</b>
	M	12/8			
	W	12/10			
	F	12/12	<b>No final exam</b>		<b>Honors Project Due</b>