

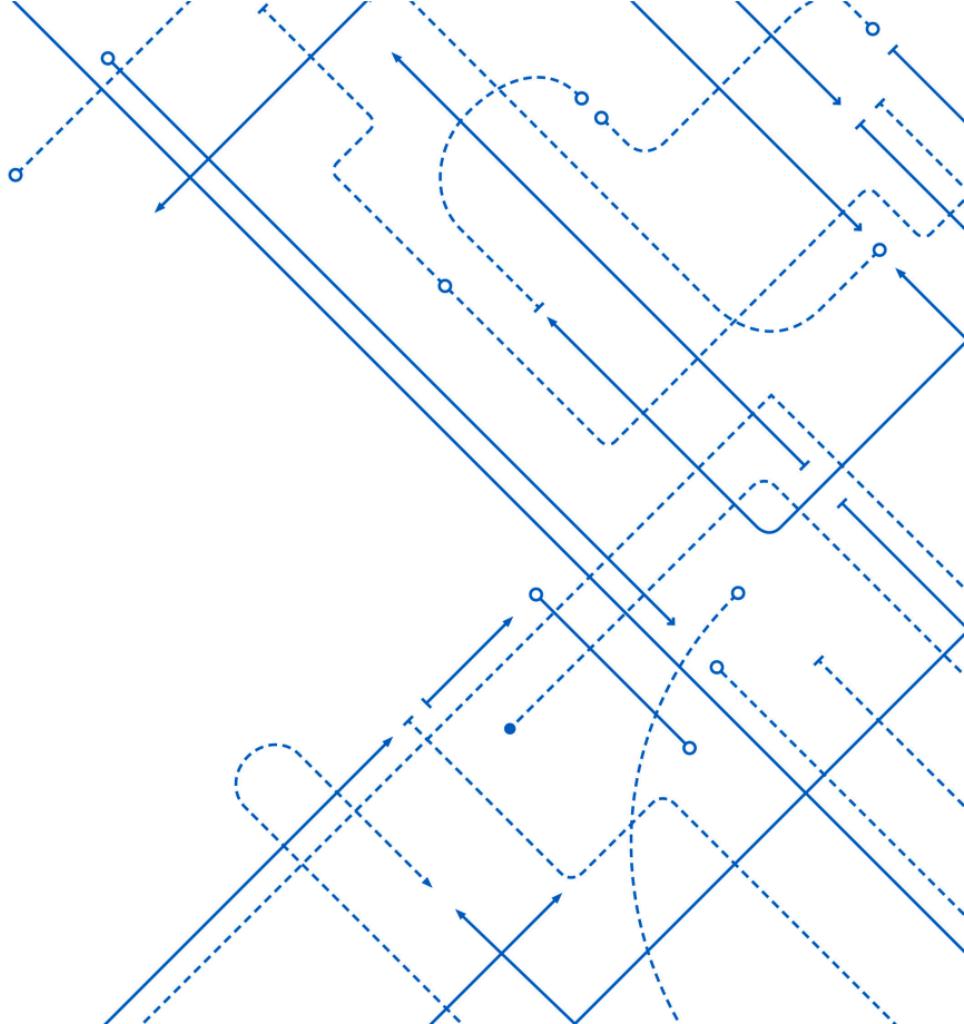
# Clustering

Kenneth (Kenny) Joseph



University at Buffalo

Department of Computer Science  
and Engineering  
School of Engineering and Applied Sciences



# From last week

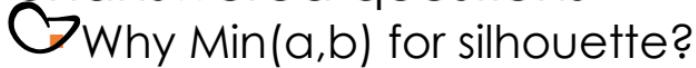


Quiz 7 review



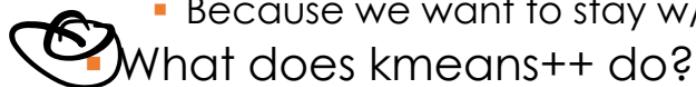
Quick Discussion about PA 3

- Unanswered questions



Why  $\text{Min}(a,b)$  for silhouette?

- Because we want to stay w/in  $[-1,1] \dots ?$



What does kmeans++ do?

→ Exam update

→ Hand off on Quiz 8 'til tonight

# Kmeans++

A slightly smarter random initialization

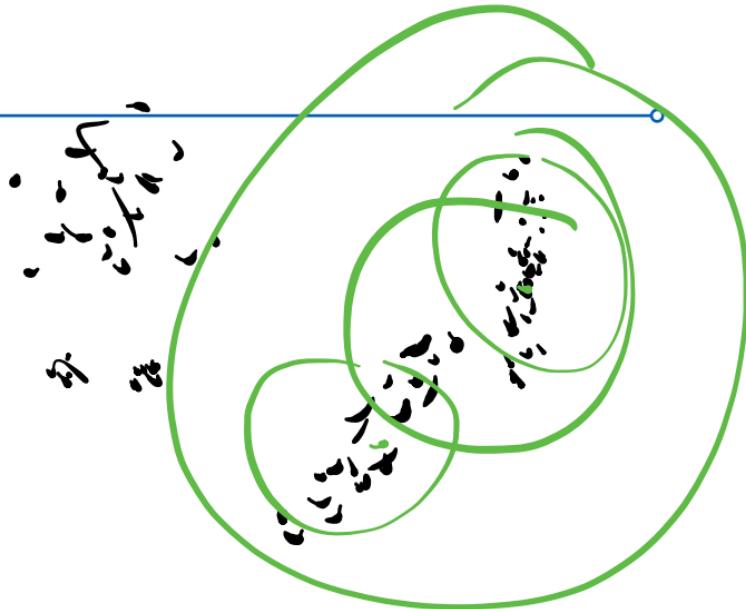
1. Choose first cluster  $\mu_1$  from the data uniformly at random
2. For the current set of centroids (starting with just  $\mu_1$ ), compute the distance between each datapoint and its closest centroid
3. Choose a new centroid from the remaining data points with probability of  $x_i$  being chosen proportional to  $d(x_i)^2$
4. Repeat 2 and 3 until we have selected  $k$  centroids

**Slide adapted from:**

<https://courses.cs.washington.edu/courses/cse416/21sp/>

# From last week

- Quiz 7 review
  - Quick Discussion about PA 3
  - Unanswered questions
    - Why  $\text{Min}(a,b)$  for silhouette?
      - Because we want to stay w/in  $[-1,1] \dots ?$
    - What does kmeans++ do?
- Review:**
- **Name one limitation of the Kmeans clustering algorithm that can be addressed using Gaussian mixture models**



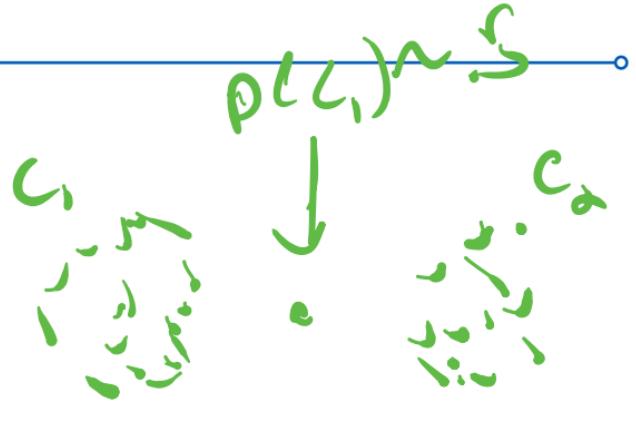
University at Buffalo

Department of Computer Science  
and Engineering

School of Engineering and Applied Sciences

# From last week

- Quiz review
- Discussion about PA 3
- Unanswered questions
  - Why  $\text{Min}(a,b)$  for silhouette?
    - Because we want to stay w/in  $[-1,1] \dots ?$
  - What does kmeans++ do?
- **Review quiz:**
  - **Name one limitation of the Kmeans clustering algorithm that can be addressed using Gaussian mixture models**
    - Clusters with different shapes/orientations
    - One we didn't discuss: soft clustering



# Today: PA4, brief intro (up later)

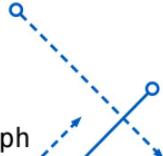
---



University at Buffalo

Department of Computer Science  
and Engineering

School of Engineering and Applied Sciences



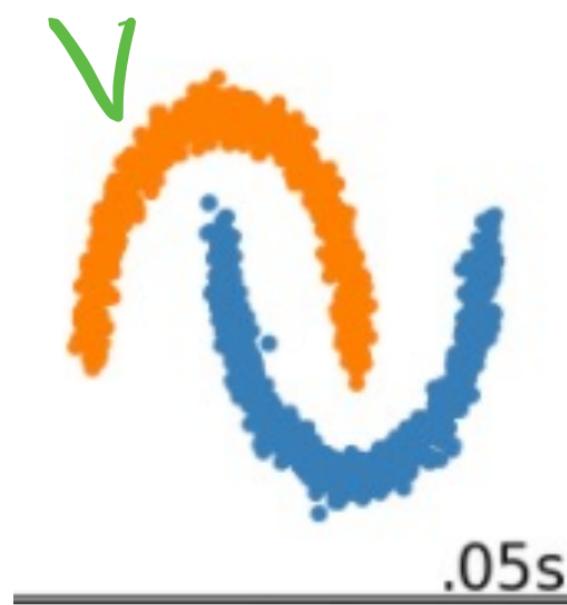
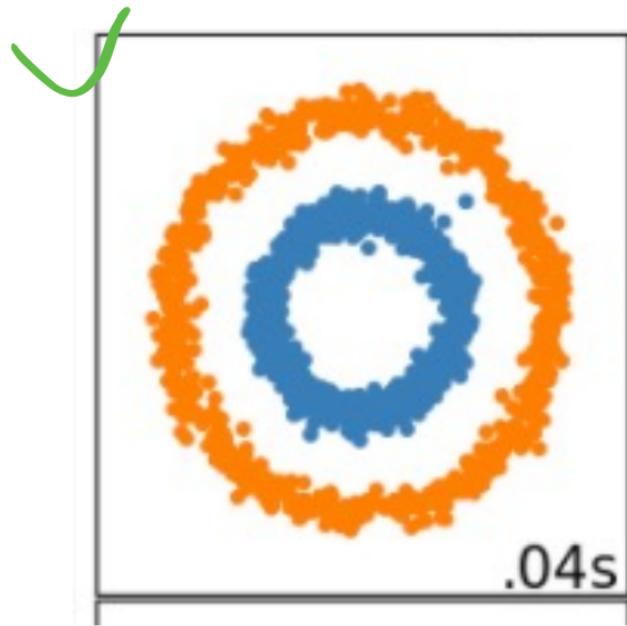
# Today: One more clustering algorithm

---

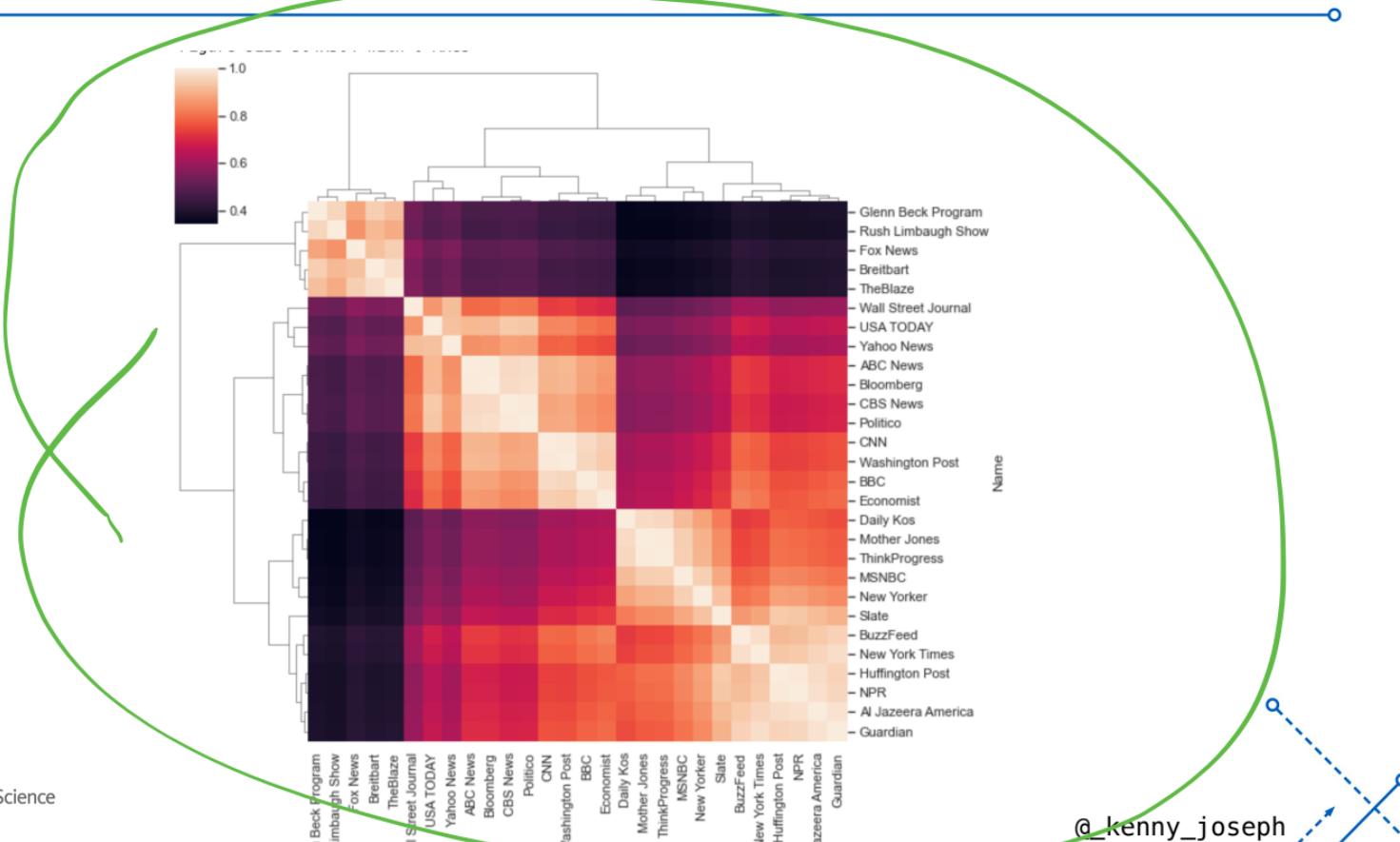
- **Hierarchical clustering**
- Manual demo, and more code



# oooohhh



# ahhhhhhhh



University at Buffalo

Department of Computer Science  
and Engineering

School of Engineering and Applied Sciences

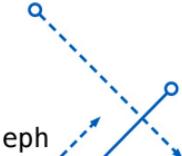
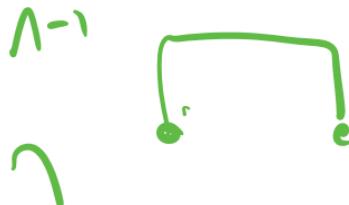
@kenny\_joseph

# The basic idea of hierarchical clustering



- Group close things together (I know, crazy, right)
- End up with a **hierarchy of clusters**
- **Quiz: How would you visualize a hierarchy?**

cluster



# The basic idea of hierarchical clustering

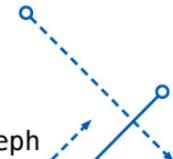
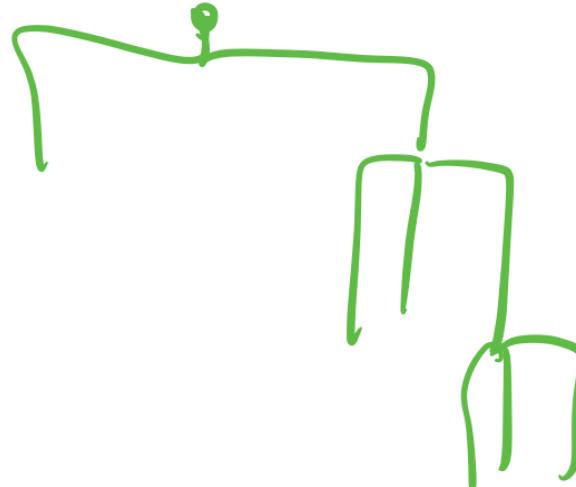
- Group close things together (I know, crazy, right)
- End up with a **hierarchy of clusters**
- **Quiz: Where do we see hierarchies?**

Species  
access to resources  
wikipedia



# The basic idea of hierarchical clustering

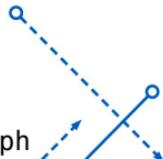
- Group close things together (I know, crazy, right)
- End up with a **hierarchy of clusters**
- **Quiz: How do you draw a hierarchy?**



# The basic idea of hierarchical clustering

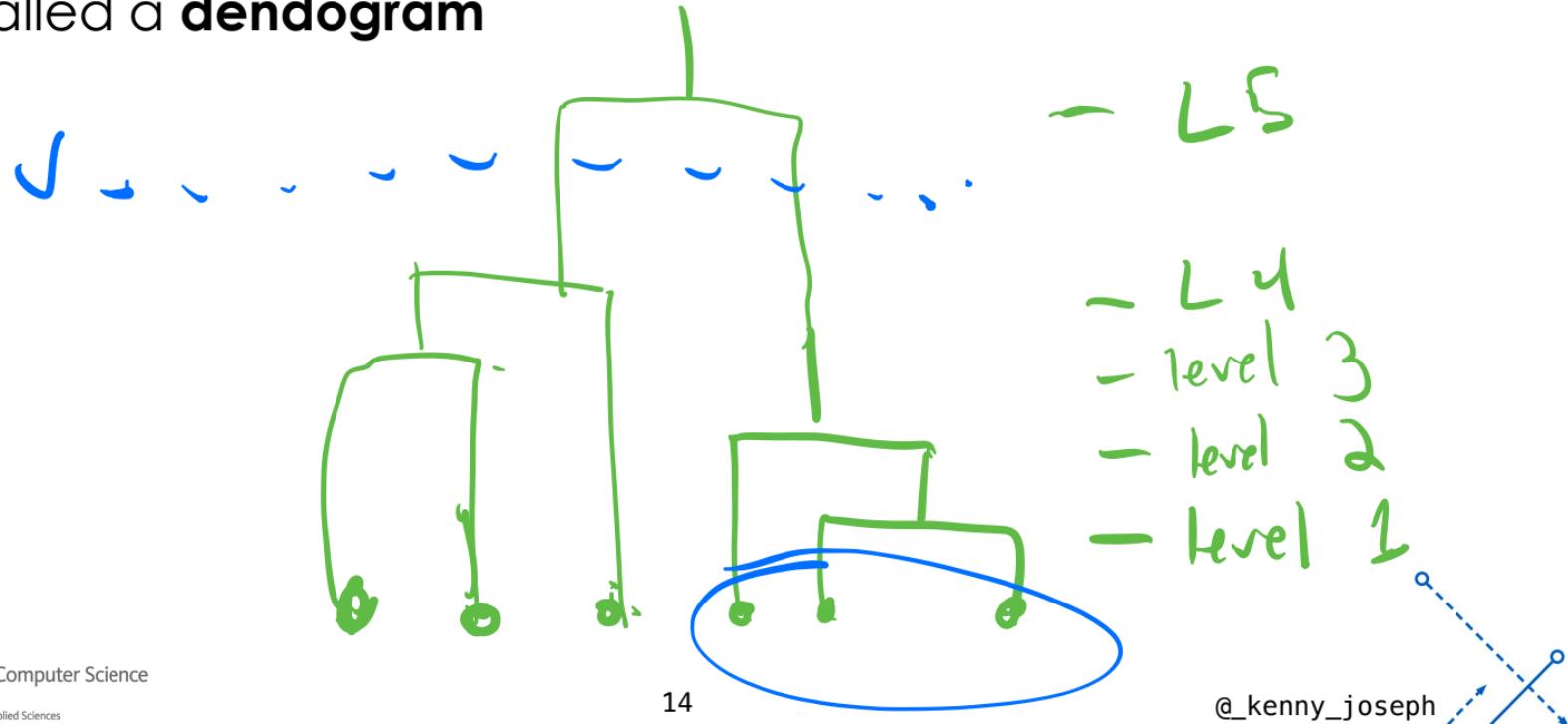
- Group close things together (I know, crazy, right)
- End up with a **hierarchy of clusters**
- **Quiz: Why might a hierarchy be useful?**

multiple levels  
of  
granularity



# Dendograms

- The intuitive drawing that we made for hierarchies is called a **dendrogram**



# Approaches to hierarchical clustering

---

- **Divisive**, a.k.a. *top-down*

- Start with all the data in one big cluster and then recursively split the data into smaller clusters

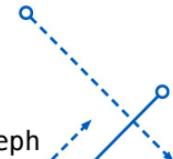


## Agglomerative, a.k.a. bottom-up:

- Start with each data point in its own cluster. Merge clusters until all points are in one big cluster.

Slide adapted from:

<https://courses.cs.washington.edu/courses/cse416/21sp/>



# Divisive Clustering – brief example

- We will focus on agglomerative clustering, but you should get the main idea of divisive clustering (start with one cluster, recursively divide it)
- Quiz: What is an algorithm you could use to do this?**



# Devisive clustering (cont.)



- For divisive clustering, you need to make the following choices:

- Which algorithm to use
- How many clusters per split
- When to split vs when to stop
  - Max cluster size**  
Number of points in cluster falls below threshold
  - Max cluster radius**  
distance to furthest point falls below threshold
  - Specified # of clusters**  
split until pre-specified # of clusters is reached

Slide adapted from:

<https://courses.cs.washington.edu/courses/cse416/21sp/>



# Agglomerative Clustering algorithm

---

- Initialize each point in its own cluster
- Define a distance metric between clusters

While there is more than one cluster

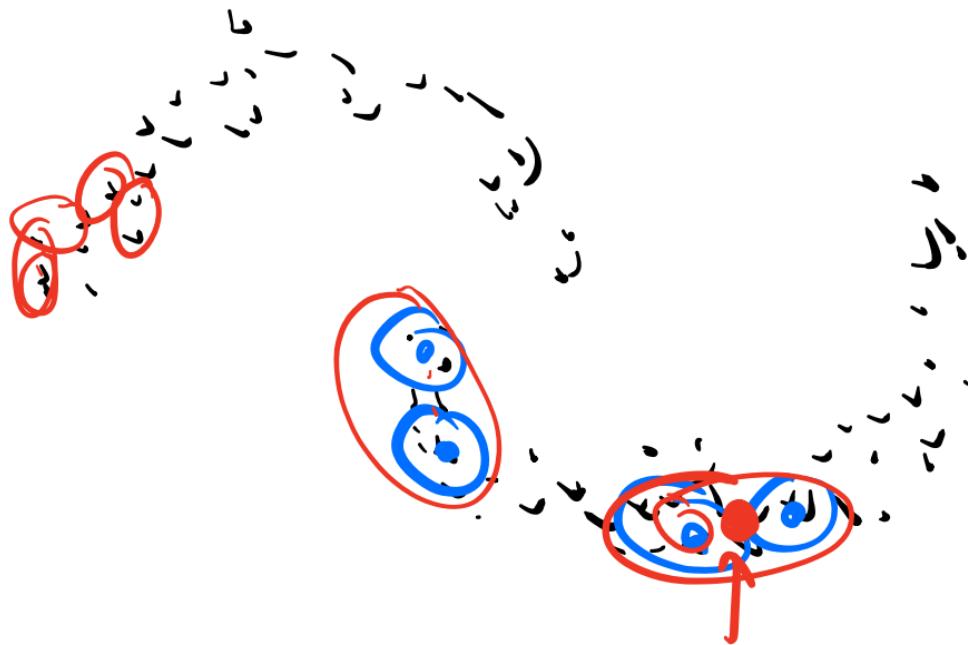
- Merge the two closest clusters



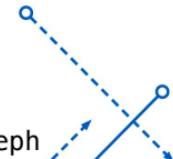
# Agglomerative Clustering worked example



# Agglomerative Clustering worked example



# Agglomerative Clustering worked example



# Agglomerative Clustering worked example



University at Buffalo

Department of Computer Science  
and Engineering

School of Engineering and Applied Sciences

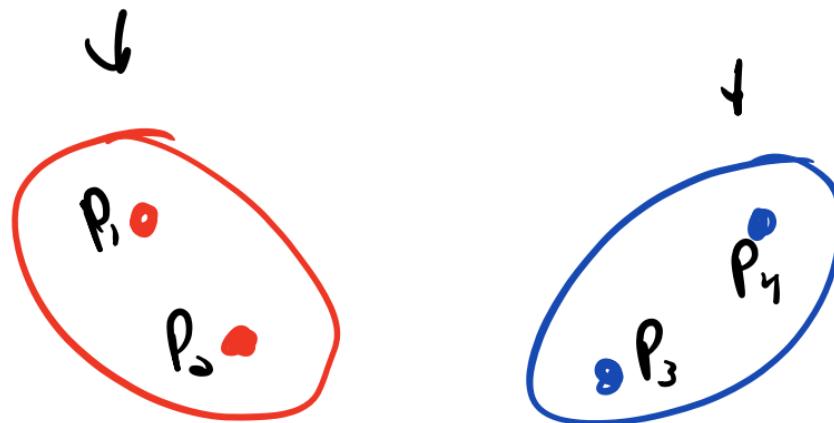
# Agglomerative Clustering (cont.)

- For agglomerative clustering, you need to make the following choices:
  1. Distance metric
  2. Linkage function
    - Single Linkage
    - Complete Linkage
    - Centroid Linkage
    - Others (Ward)
  3. Where and how to cut dendrogram

**Slide adapted from:**

<https://courses.cs.washington.edu/courses/cse416/21sp/>

# Single Linkage



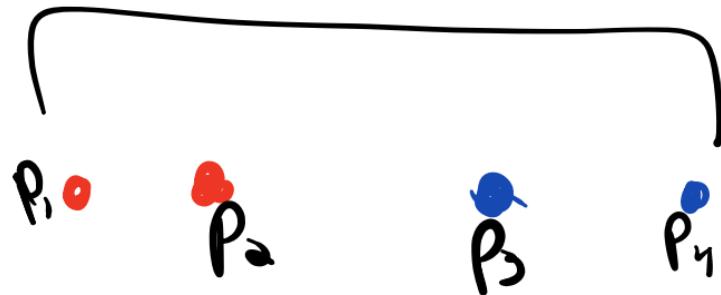
$$\min d(x_i, x_j)$$

which pairs  
would I  
use to  
define single  
linkage?

$P_3 \text{ & } P_4$

@kenny\_joseph

# Complete Linkage



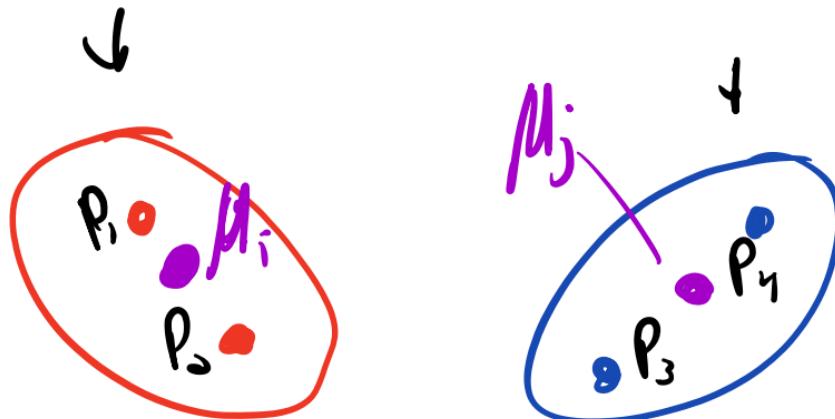
$$\max d(x_i, x_j)$$

which pairs  
would I  
use to  
define complete  
linkage?

$p_1$  &  $p_4$

@kenny\_joseph

# Centroid Linkage



$d \rightarrow$  Euclidean  
or  
 $d \rightarrow$  Manhattan

$$\min d(\mu_i, \mu_j)$$



University at Buffalo

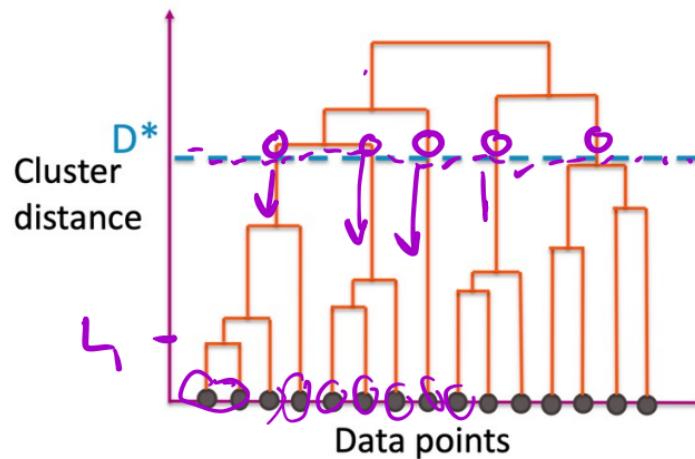
Department of Computer Science

and Engineering

School of Engineering and Applied Sciences

# Agglomerative Clustering (quiz)

How many clusters would we have if we use this threshold?



Slide adapted from:

<https://courses.cs.washington.edu/courses/cse416/21sp/>

# Code demo/think-through

---



University at Buffalo

Department of Computer Science  
and Engineering

School of Engineering and Applied Sciences

