Unsupervised Learning

Danna Gurari
University of Texas at Austin
Spring 2020

https://www.ischool.utexas.edu/~dannag/Courses/IntroToMachineLearning/CourseContent.html

Review

- Last week:
  - Machine Learning for Sequential Data
  - Recurrent Neural Networks (RNNs)
  - Training Deep Neural Networks: Hardware & Software

- Assignments (Canvas):
  - Project proposal due yesterday
  - Project outline due next week
  - Prototype of final project system due next week

- Questions?

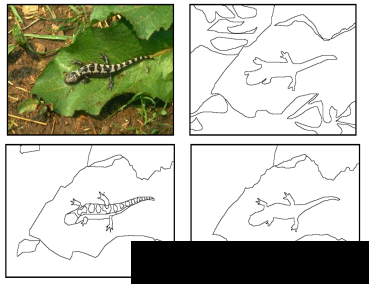
Today’s Topics

- Machine Learning for Unlabeled Data
- Autoencoders
- Clustering
- Guest: Dr. Suyog Jain from PathAI

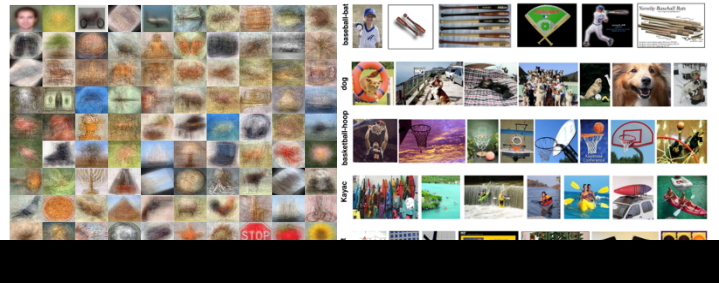
Today’s Topics

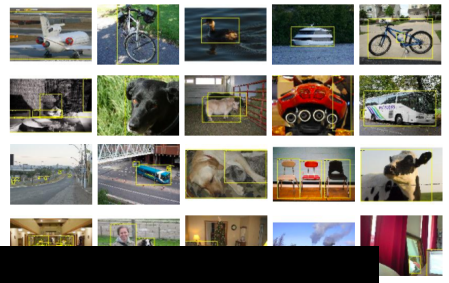
- Machine Learning for Unlabeled Data
- Autoencoders
- Clustering
- Guest: Dr. Suyog Jain from PathAI

How Have Machines Learned So Far in this Class?



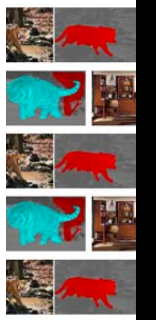
BSD



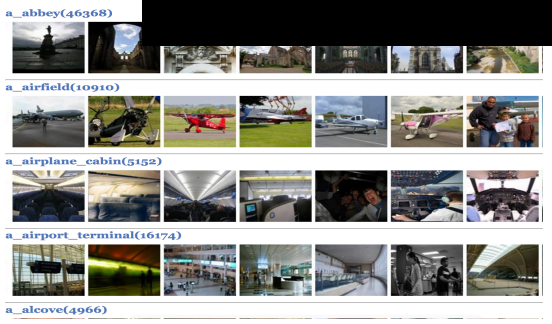


2)

Large labelled datasets

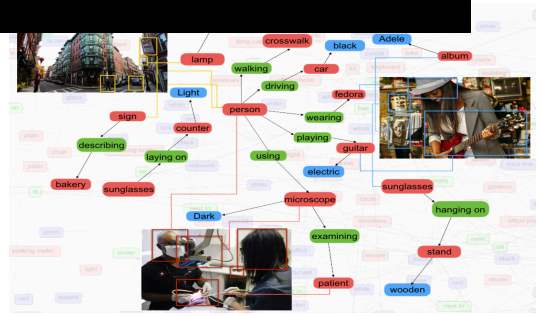


Labe



Places (2014)

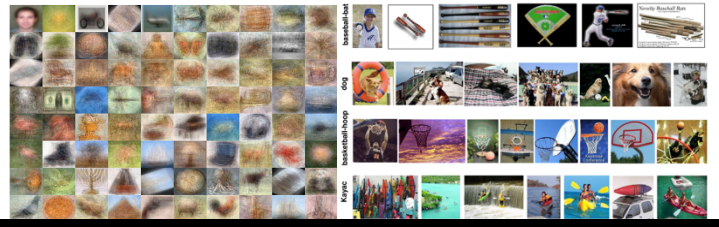
MS COCO (2014)

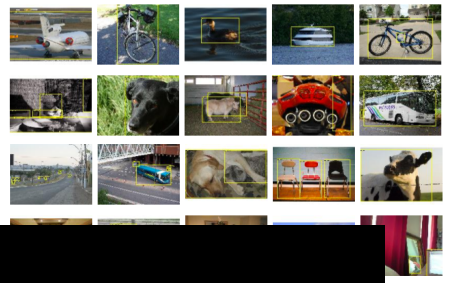


Visual Genome (2016)

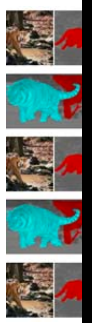
Slide Credit: http://vision.cs.utexas.edu/slides/mit-ibm-august2018.pdf

Why Not Rely On Large Labelled Datasets?

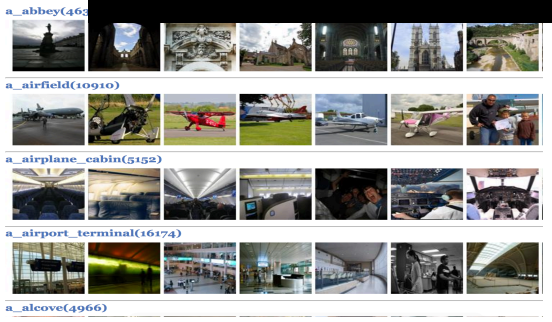




- Expensive
- Relatively Slow
- Disconnect from Human Learning

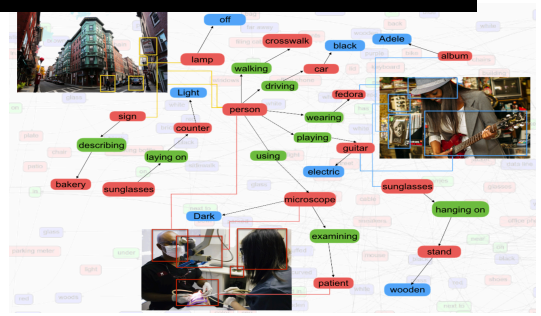


La



Places (2014)

MS COCO (2014)



Visual Genome (2016)

Slide Credit: http://vision.cs.utexas.edu/slides/mit-ibm-august2018.pdf

Intuition: How Do Humans Learn?

With Supervision

No Supervision

https://pixabay.com/en/toddler-learning-book-child-423227/

https://www.maxpixel.net/Father-Child-Family-Dad-Baby-Daughter-3046495

Intuition: How Do Humans Learn?

- Experience (Unsupervised):
  - Idea: learn to group objects into one class by seeing many of them

- Supervision:
  - Idea: learn to group objects into one class because someone tells us to

Recall: Types of Learning Tasks

- Unsupervised
  - No label given for training data

- Supervised
  - Label given for training data: e.g., “cat”

Recall: Types of Learning Tasks

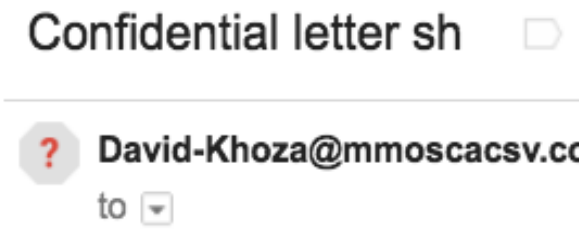
- Unsupervised
  - No label given for training data

- Supervised
  - Label given for training data: e.g., “berimbau”

Recall: Types of Learning Tasks

- Unsupervised
  - No label given for training data

- Supervised
  - Label given for training data: e.g., “yes”



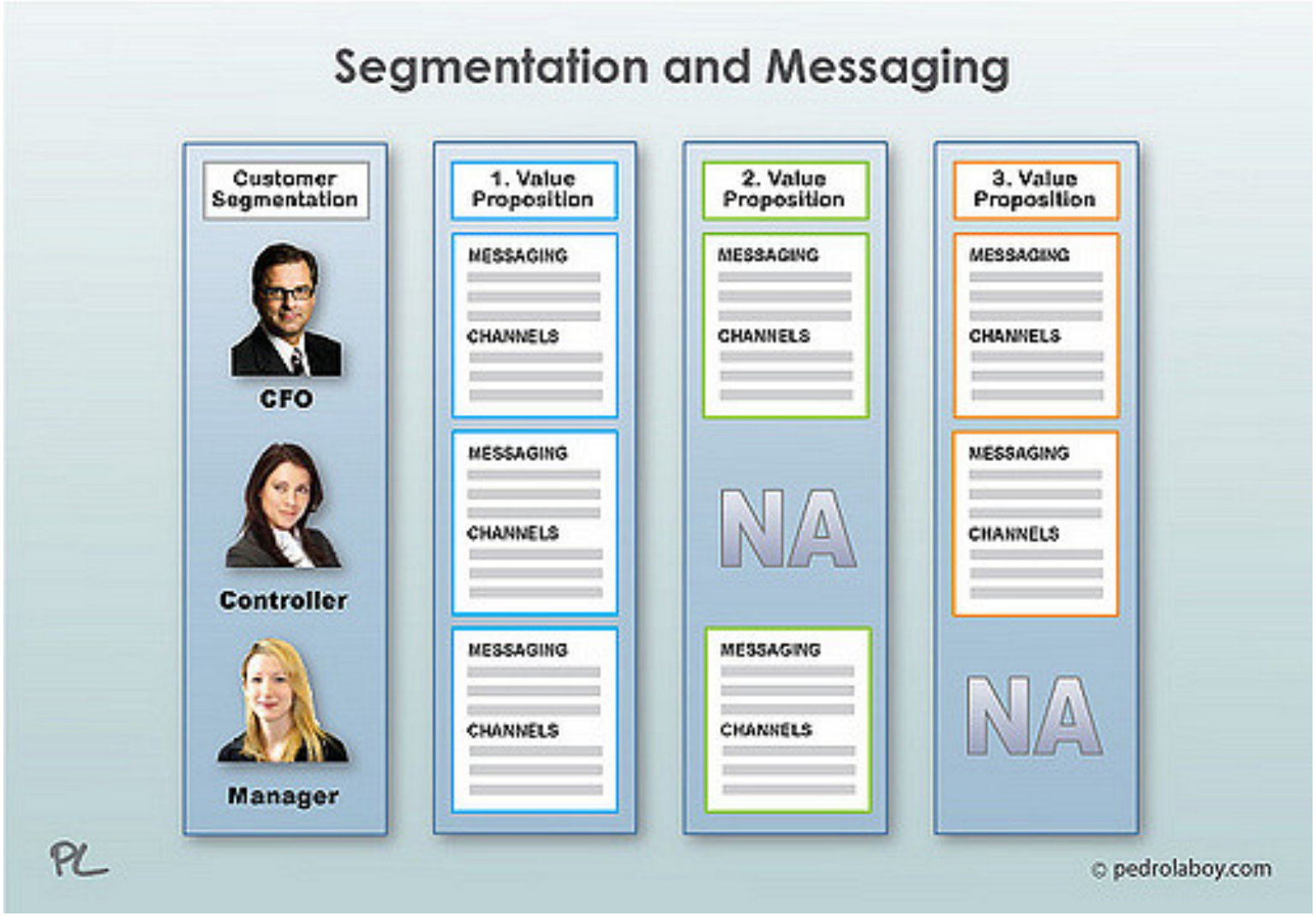
Is this email spam?

Goal: Learn from Experience To Organize Data

https://pixabay.com/en/toddler-learning-book-child-423227/

https://www.maxpixel.net/Father-Child-Family-Dad-Baby-Daughter-3046495

Real-World Applications: Customer Segmentation

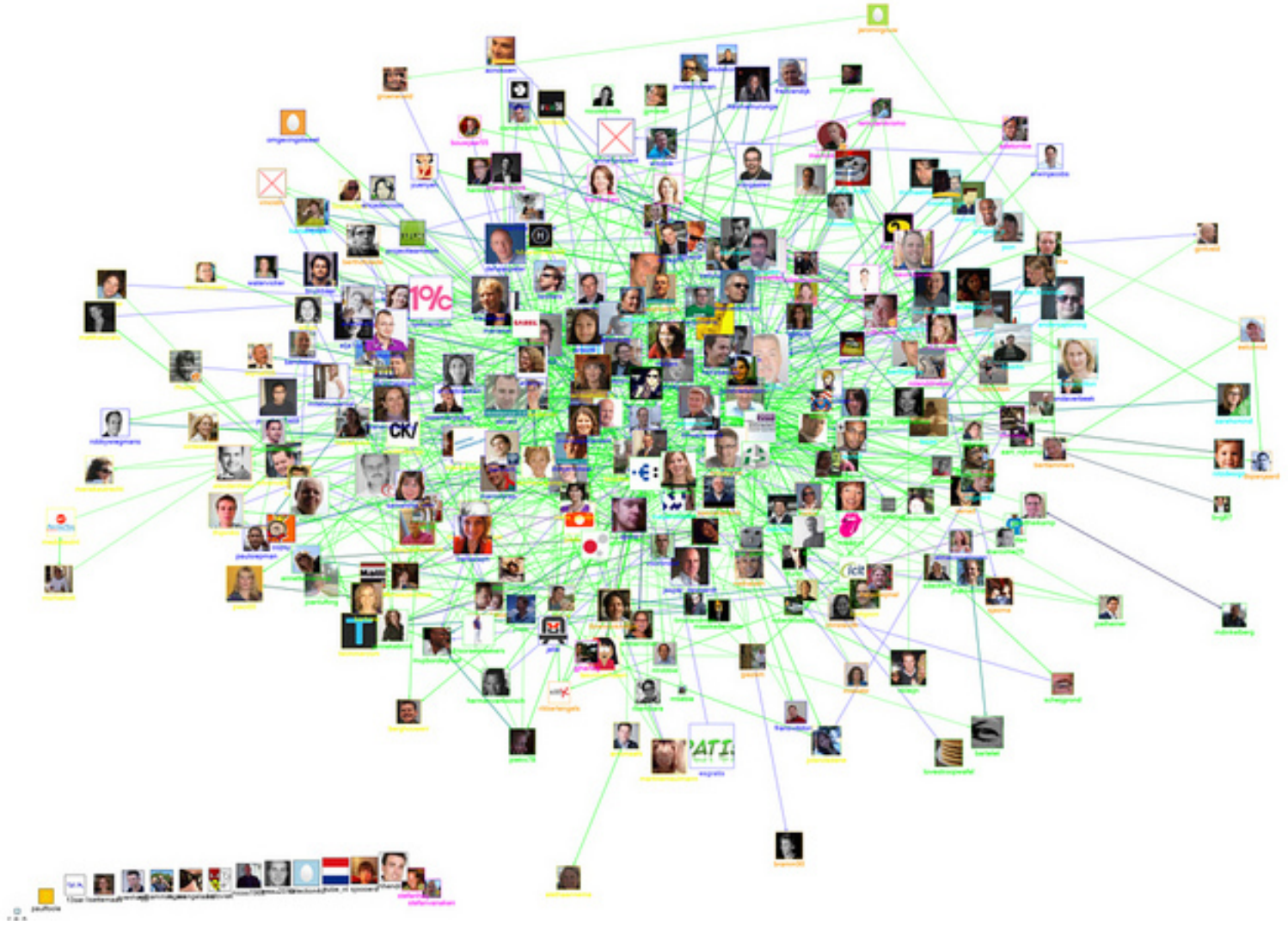


https://www.flickr.com/photos/42565140@N04/3923873188/

Real-World Applications: Recommendations

https://medium.com/@navdeepsingh_2336/scala-machine-learning-projects-recommendation-systems-d41d9eebbb06

Real-World Applications: Social Network Analysis



https://www.flickr.com/photos/marc_smith/5529685600

Real-World Applications: Fraud Detection

https://www.lejeune.marines.mil/News/Article/511667/protect-yourself-from-credit-card-fraud/

Breakout Discussion: Real-World Applications

What are other possible applications for using unsupervised learning?

Today's Topics

- Machine Learning for Unlabeled Data
- Autoencoders
- Clustering
- Guest: Dr. Suyog Jain from PathAI

Autoencoder Architecture

- Learn to copy the input to the output

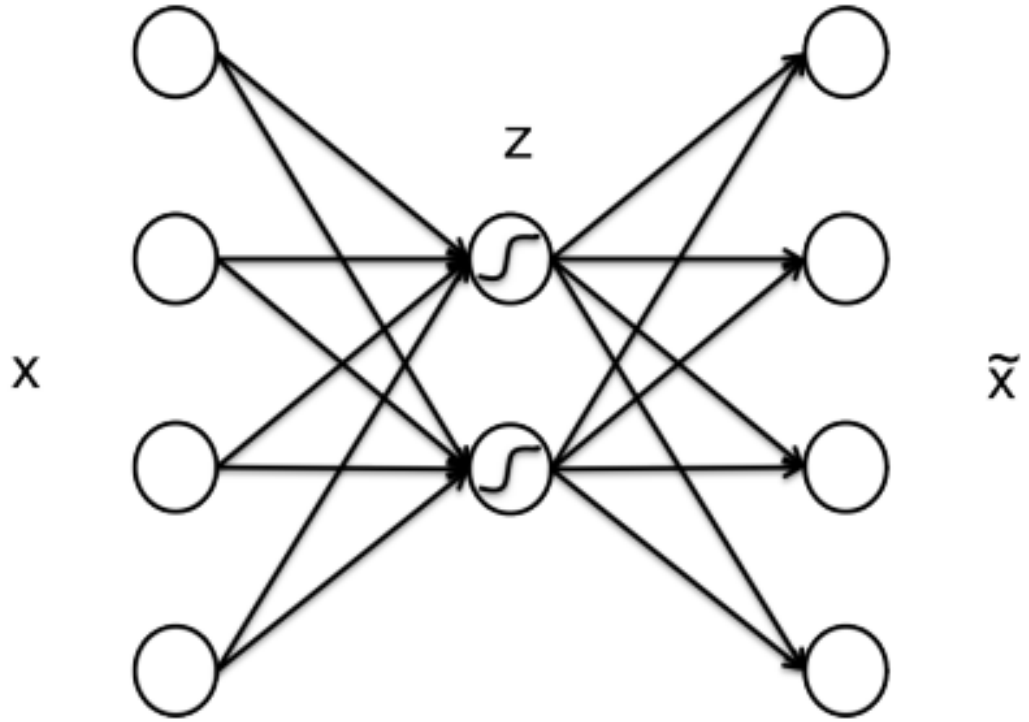


Figure Credit: https://lazyprogrammer.me/a-tutorial-on-autoencoders/

Autoencoder Architecture

- Consists of two parts:
  - **Encoder**: compresses inputs to an internal representation
  - **Decoder**: tries to reconstruct the input from the internal representation

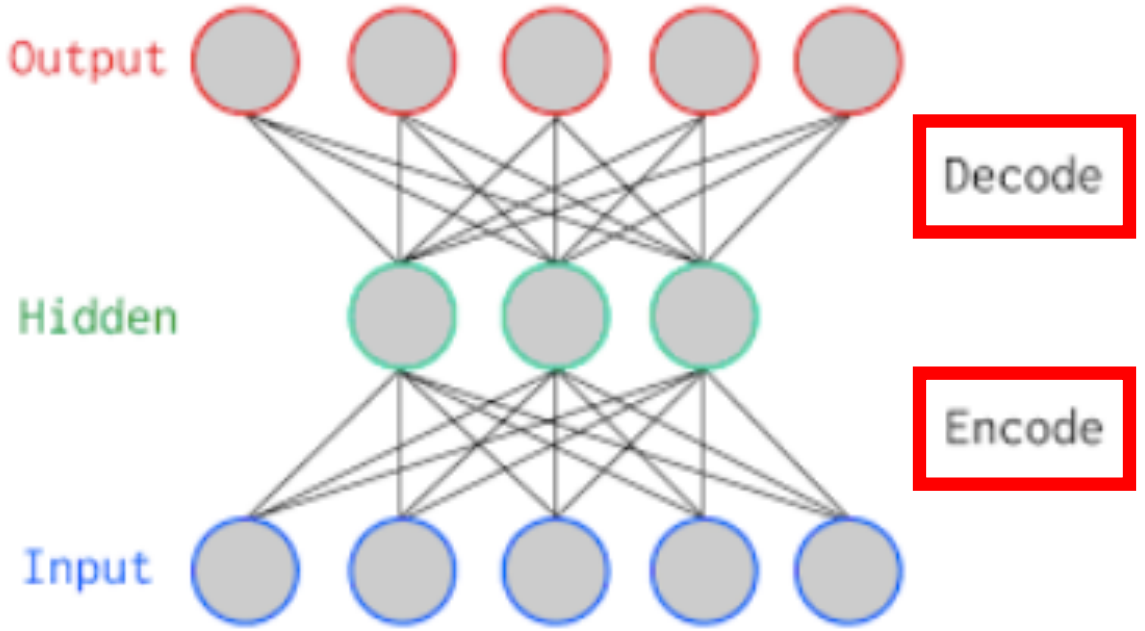
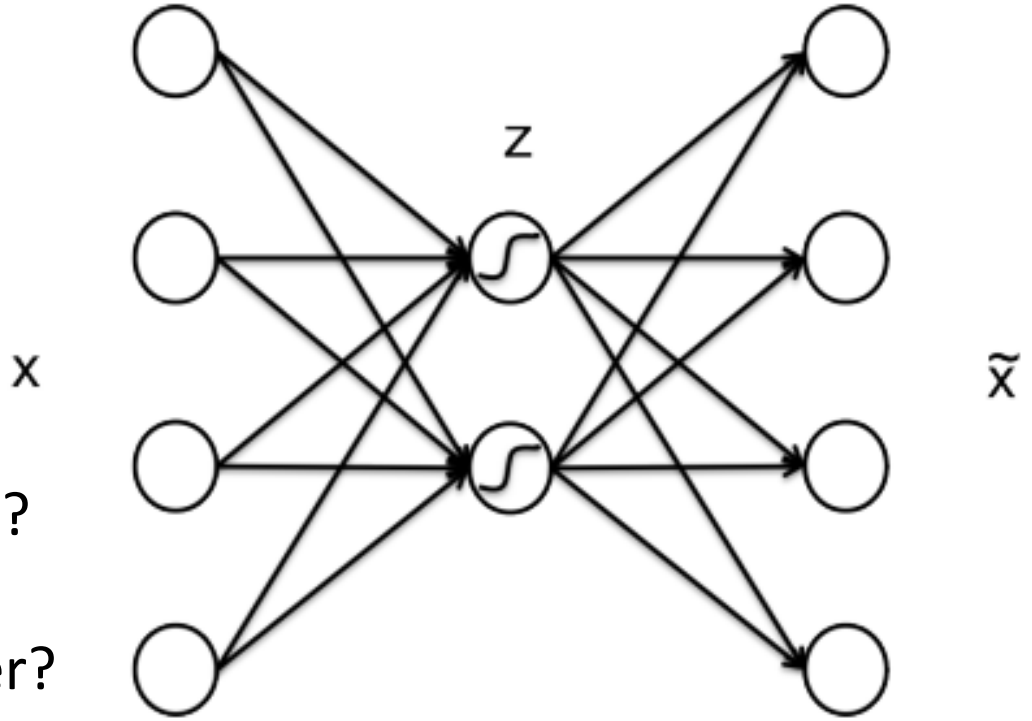


Figure Credit: https://www.datacamp.com/community/tutorials/autoencoder-keras-tutorial

Autoencoder Architecture

- Given this input 620 x 426 image (264,120 pixels):



  - What would a perfect autoencoder predict?
    - Itself
  - What number of nodes are in the final layer?
    - 264,120

Figure Credit: https://lazyprogrammer.me/a-tutorial-on-autoencoders/

Autoencoder Training

How do you train a neural network?

Autoencoder Training

Repeat until stopping criterion met:

1. **Forward pass**: propagate training data through network to make prediction
2. **Backward pass**: using predicted output, calculate error gradients backward
3. Update each weight using calculated gradients

Autoencoder

What are useful applications for autoencoders?

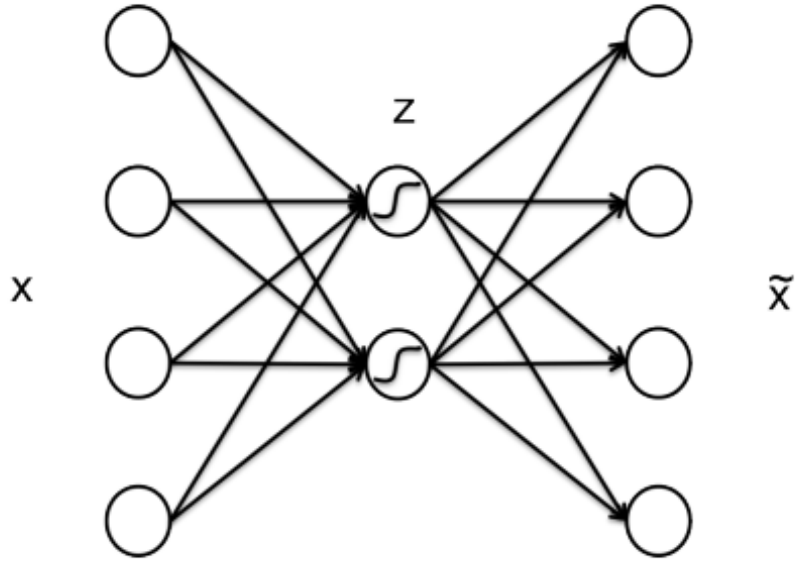


Figure Credit: https://lazyprogrammer.me/a-tutorial-on-autoencoders/

Autoencoders: Dimensionality Reduction

- Intuition: which number sequence is easier to remember?
  - **A:** 30, 27, 22, 11, 6, 8, 7, 2
  - **B:** 30, 15, 46, 23, 70, 35, 106, 53, 160, 80, 40, 20, 10, 5
- **B:** need learn only two rules
  - If even, divide by 2
  - If odd, multiply by 3 and add 1

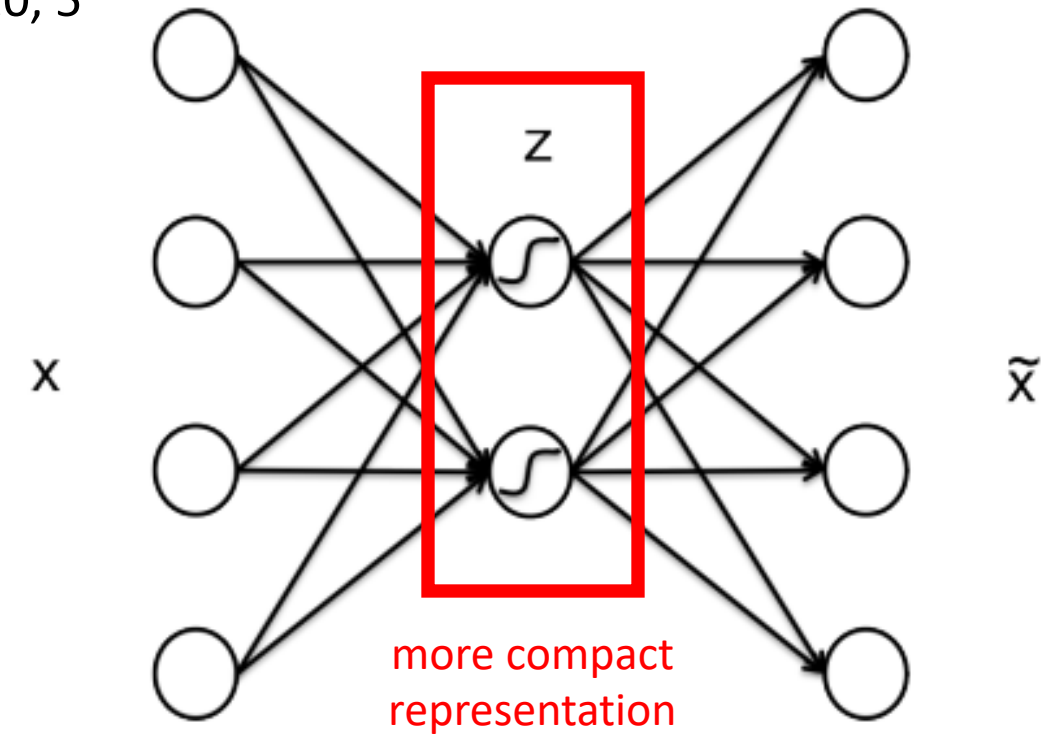
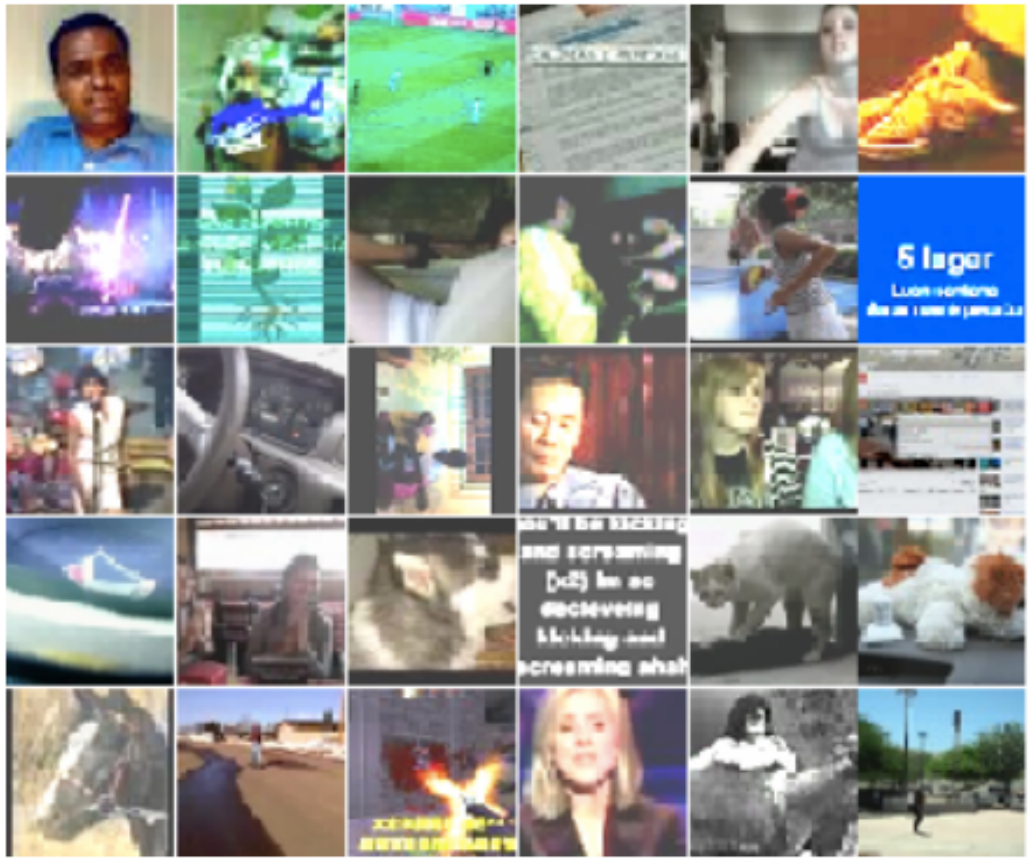


Figure Credit: https://lazyprogrammer.me/a-tutorial-on-autoencoders/

Autoencoders: Feature Extraction

- e.g., training data:
  - 1 image taken from 10 million YouTube videos
  - Each image is in color and 200x200 pixels

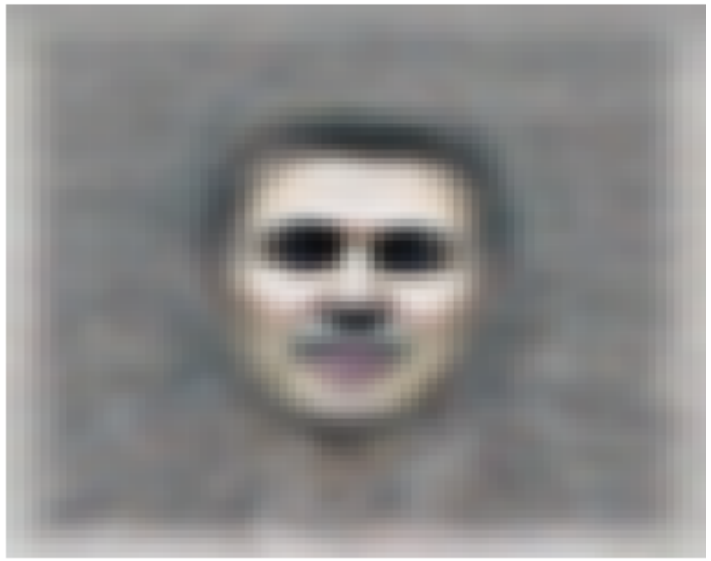


- What features do you think it learned?

Quoc V. Le et al., Building High-level Features Using Large Scale Unsupervised Learning; 2013.

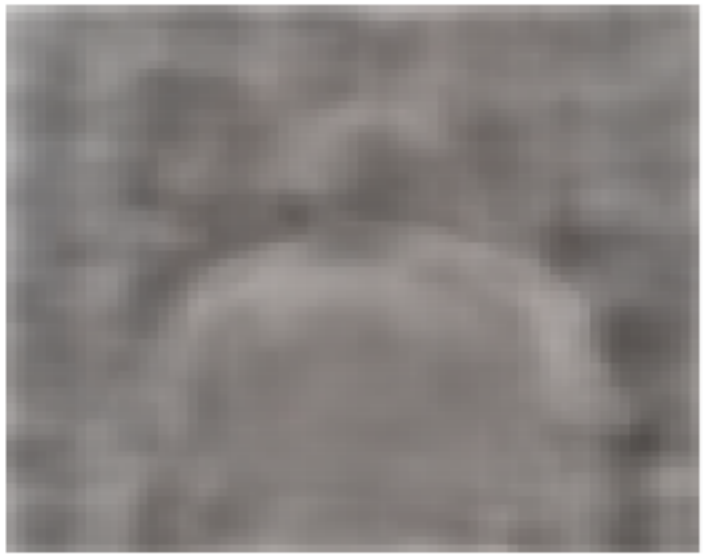
Autoencoders: Feature Extraction

- e.g., features learned include:



human face

cat face



human body

Quoc V. Le et al., Building High-level Features Using Large Scale Unsupervised Learning; 2013.

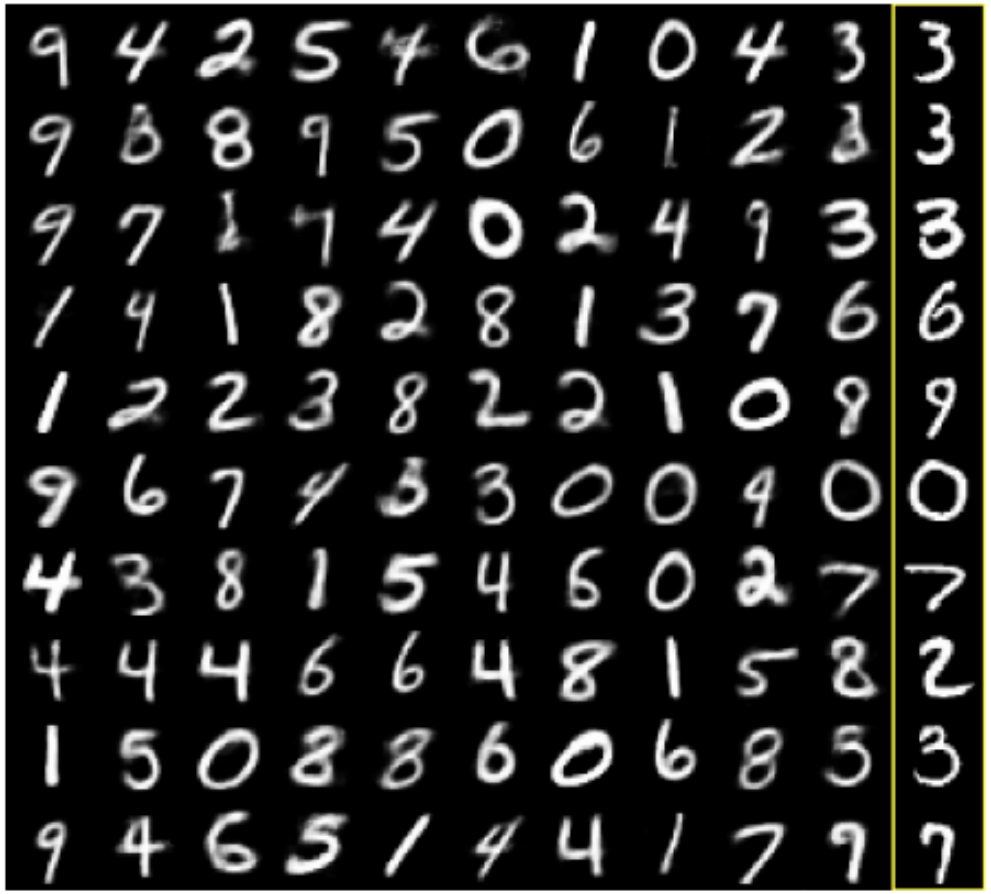
Autoencoders: Unsupervised Pretraining

- Why use unsupervised pretraining?
  - Little training data is available
  - Too costly and slow to collect labels for exclusive supervised training

- e.g., add layer after highest layer of pretrained autoencoder network (fine-tuning)

Quoc V. Le et al., Building High-level FeaturesUsing Large Scale Unsupervised Learning; 2013.

Autoencoders: Generative Models



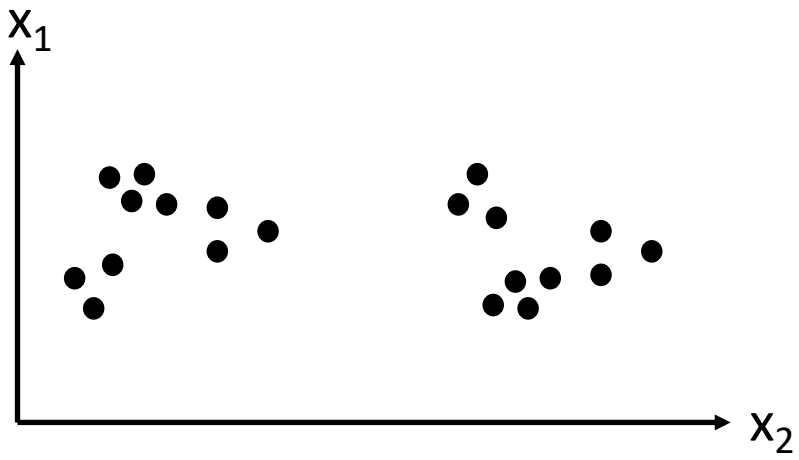
Ian Goodfellow et al; Adversarial Autoencoders; 2016

Today's Topics

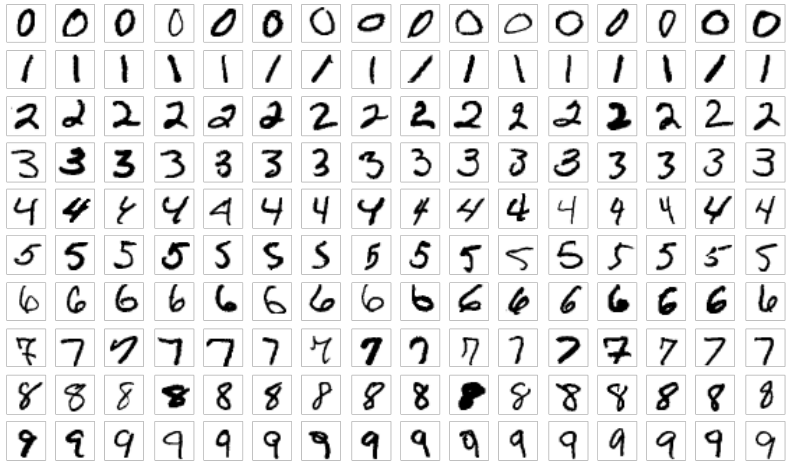
- Machine Learning for Unlabeled Data
- Autoencoders
- Clustering
- Guest: Dr. Suyog Jain from PathAI

Clustering

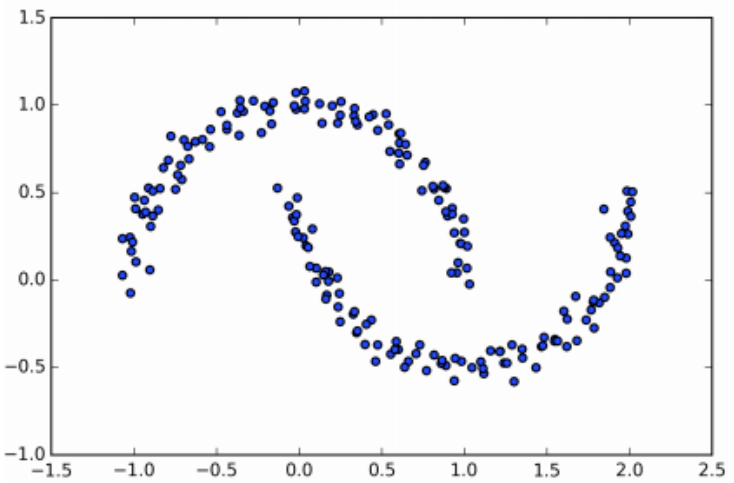
A.



B.



C.

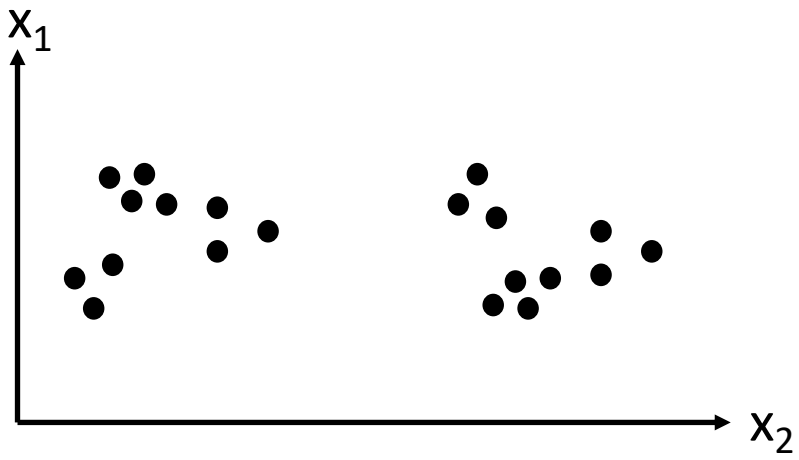


Find groupings such that entities in a group will be similar to each another and different from the entities in other groups.

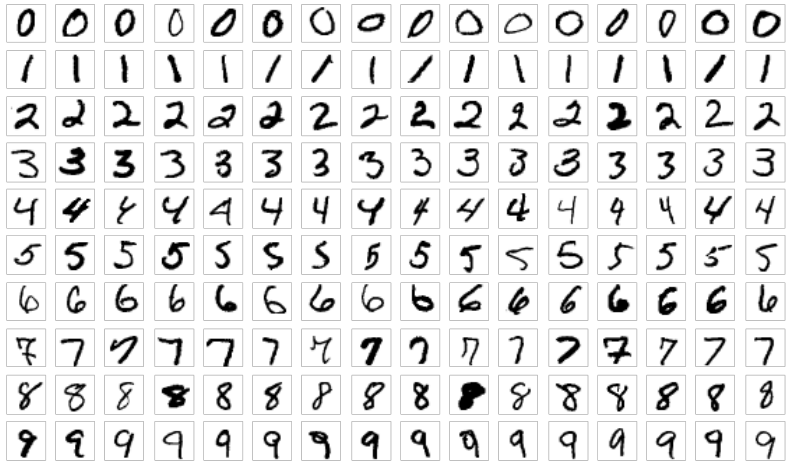
Raschka and Mirjalili; Python Machine Learning

Clustering: Key Questions

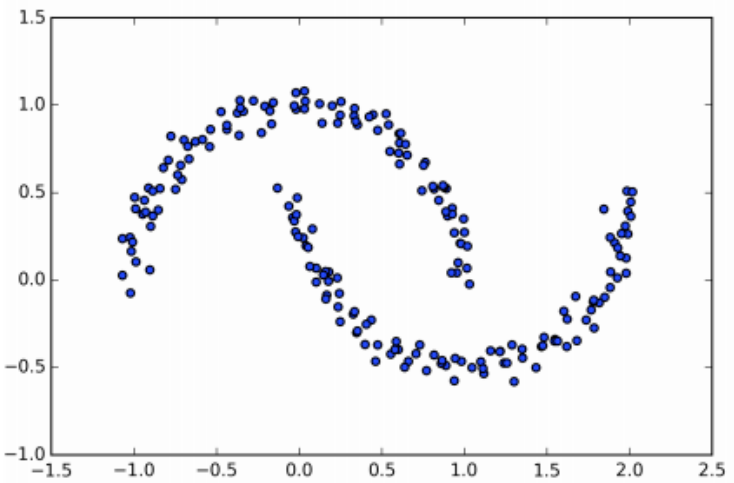
A.



B.



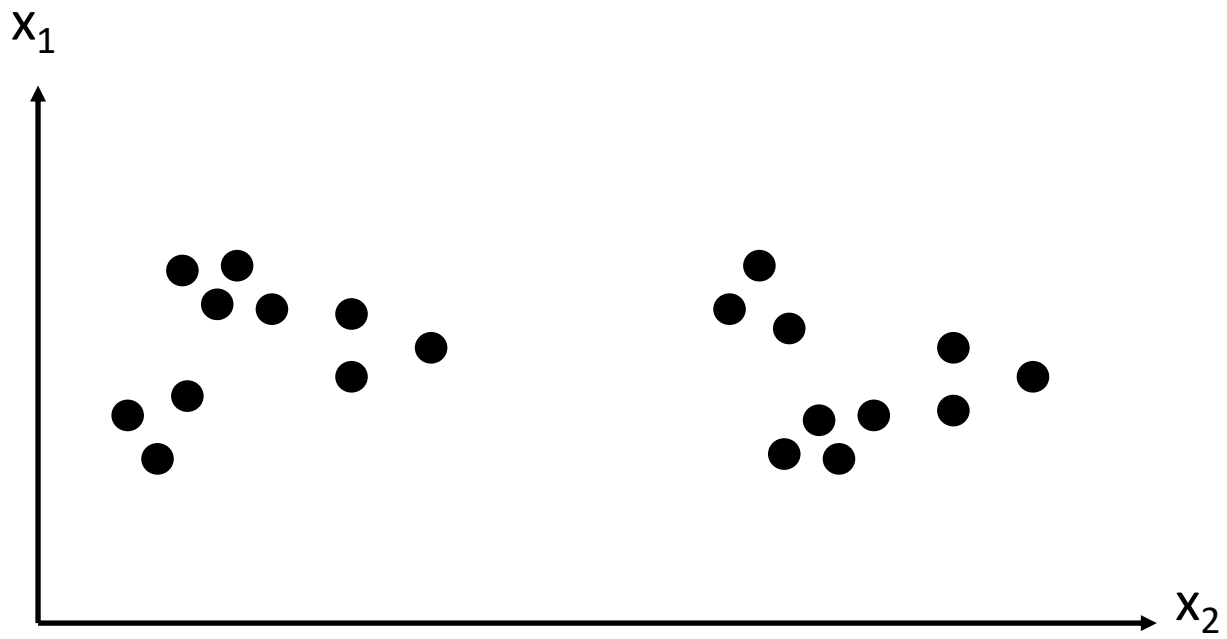
C.



- How many data clusters to create?
- What "algorithm" to use to partition the data?

Raschka and Mirjalili; Python Machine Learning

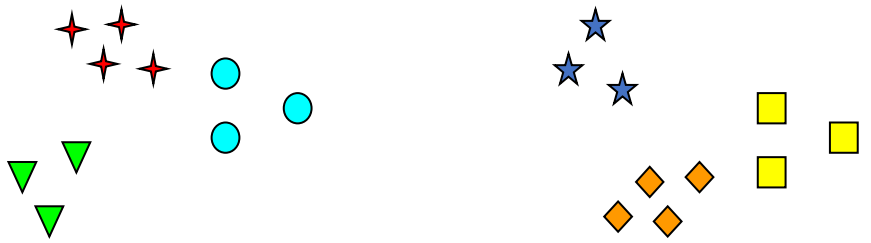
Breakout Discussion



- How many data clusters to create?
- What “algorithm” to use to partition the data?

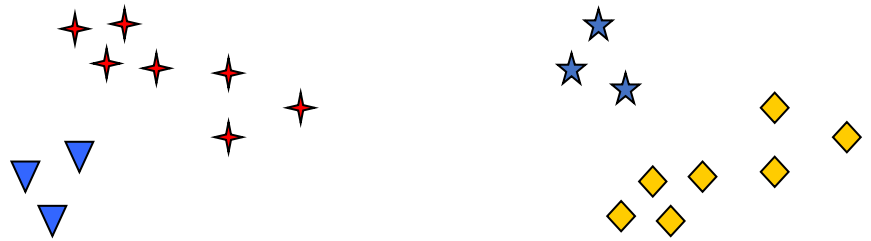
Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

How Many Clusters?



Six Clusters

Two Clusters



Four Clusters

Number of clusters can be ambiguous.

Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

Types of Clustering

- Partitional Clustering
  - A division of data objects into non-overlapping subsets (clusters) such that each data object is in exactly one subset

- Hierarchical clustering
  - A set of nested clusters organized as a hierarchical tree

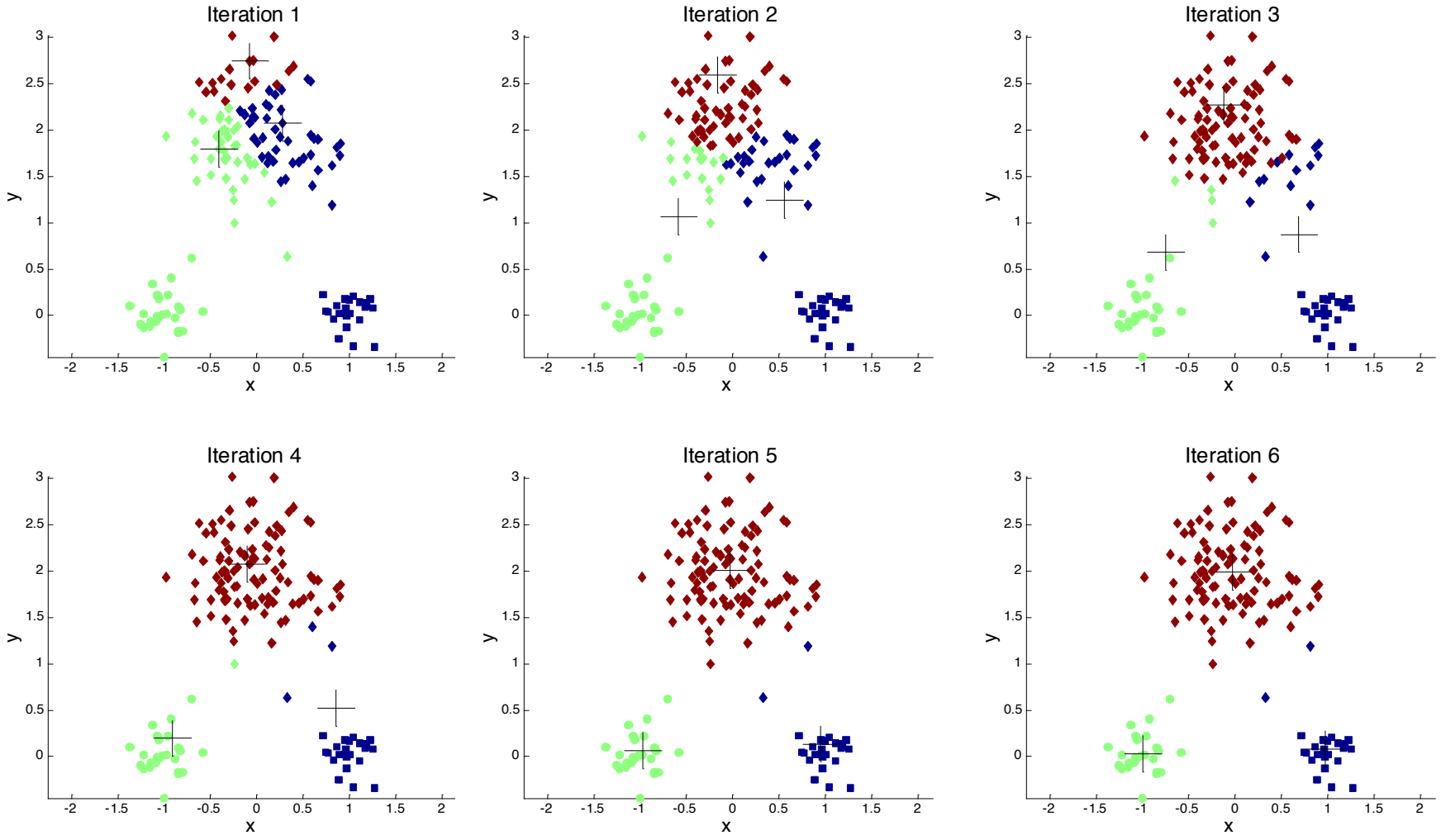
Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

K-Means Clustering

1: Select K points as the initial centroids.
2: **repeat**
3: Form K clusters by assigning all points to the closest centroid.
4: Recompute the centroid of each cluster.
5: **until** The centroids don't change

Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

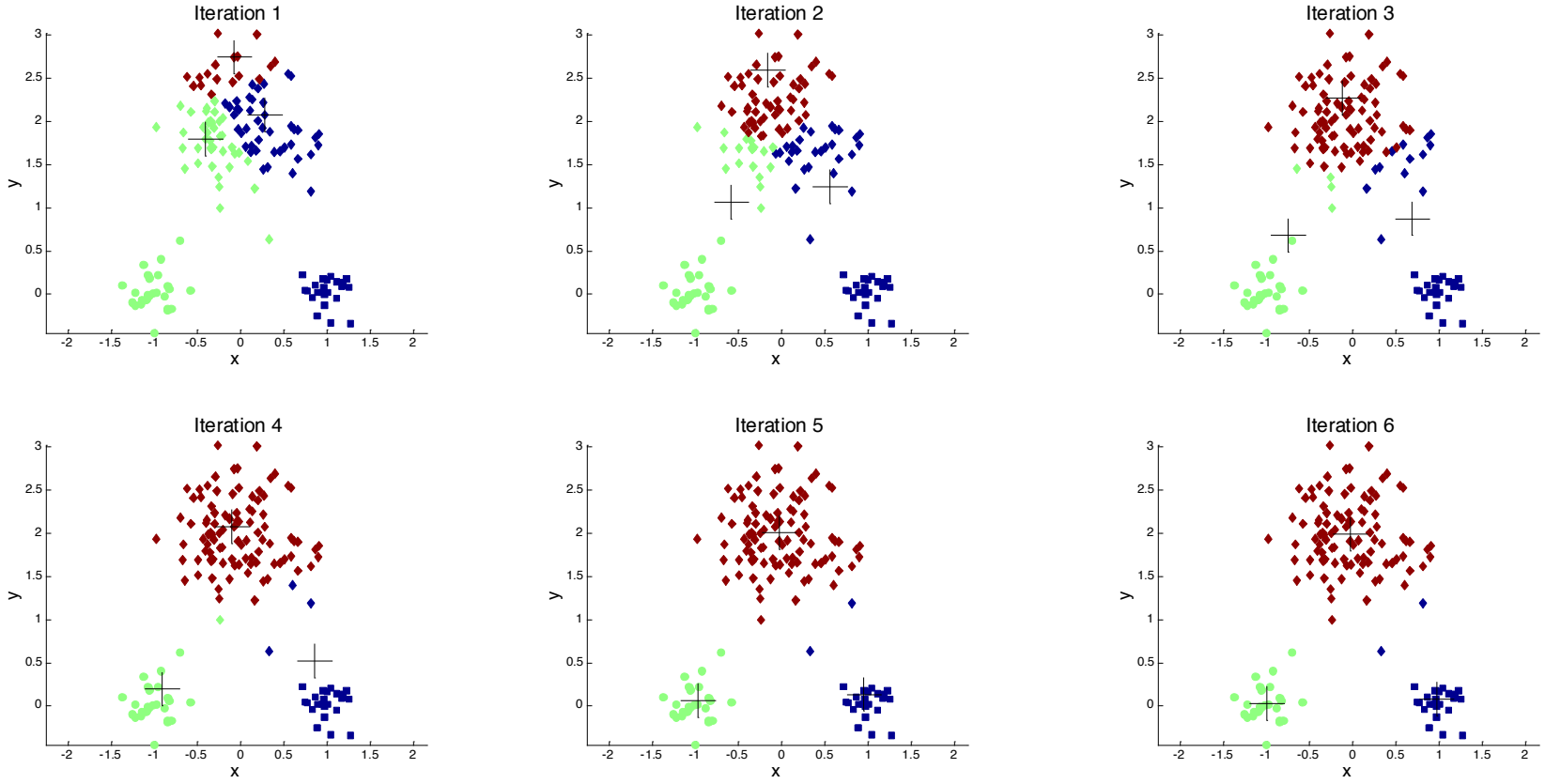
K-Means Clustering



Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

K-Means Clustering: Weaknesses?

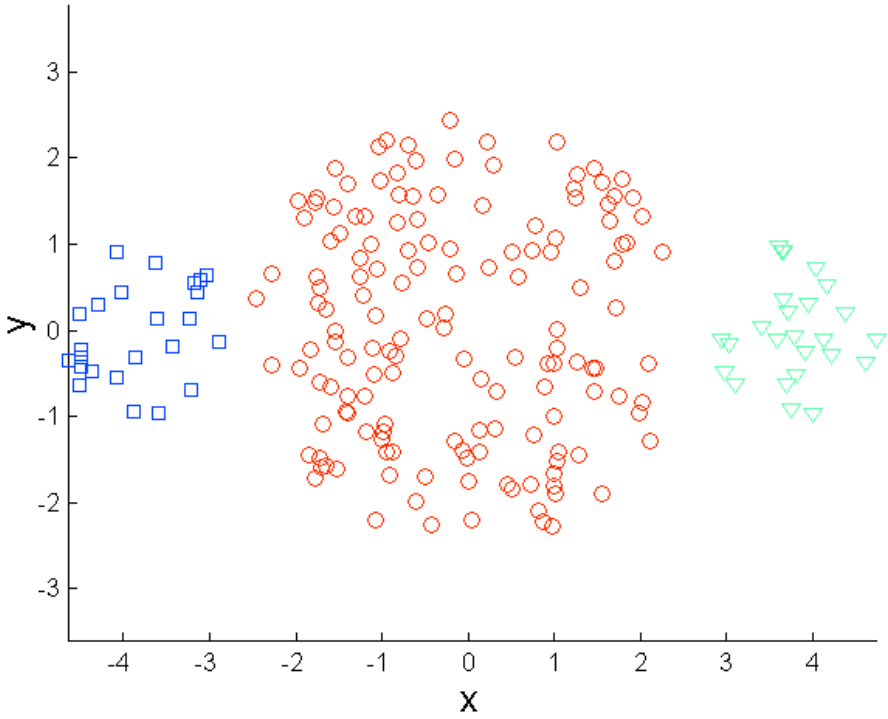
- Sensitive to initial centroids: different outcomes for same data



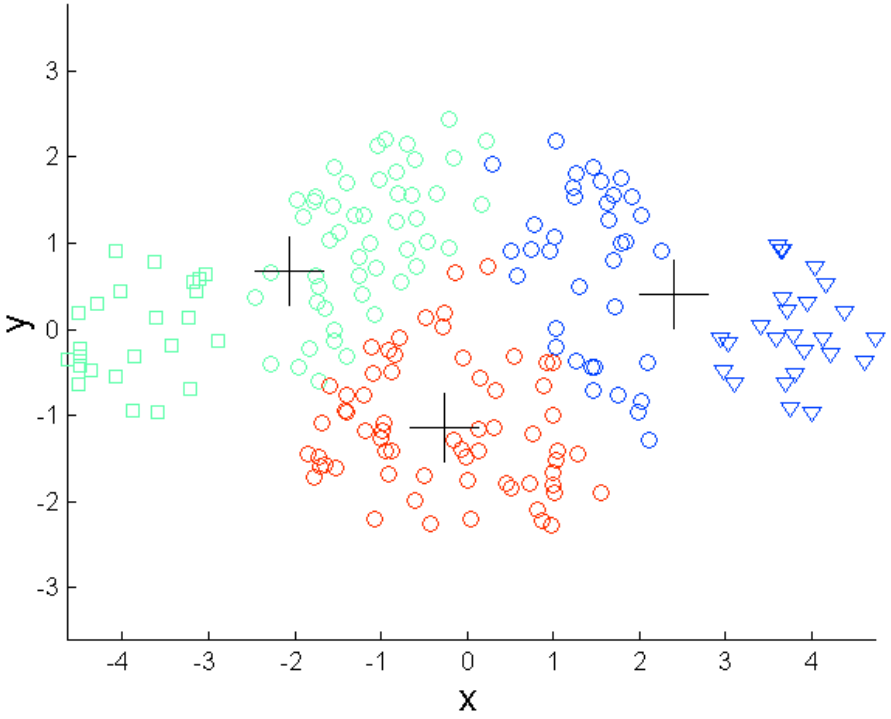
Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

K-Means Clustering: Weaknesses?

- Not robust when clusters have different sizes:



Original Points

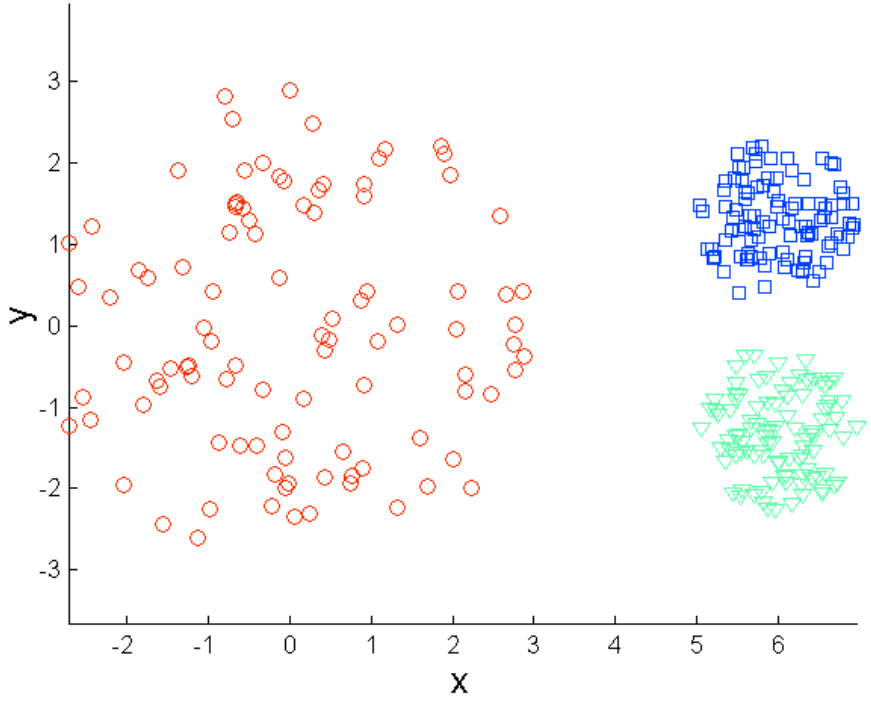


K-means (3 Clusters)

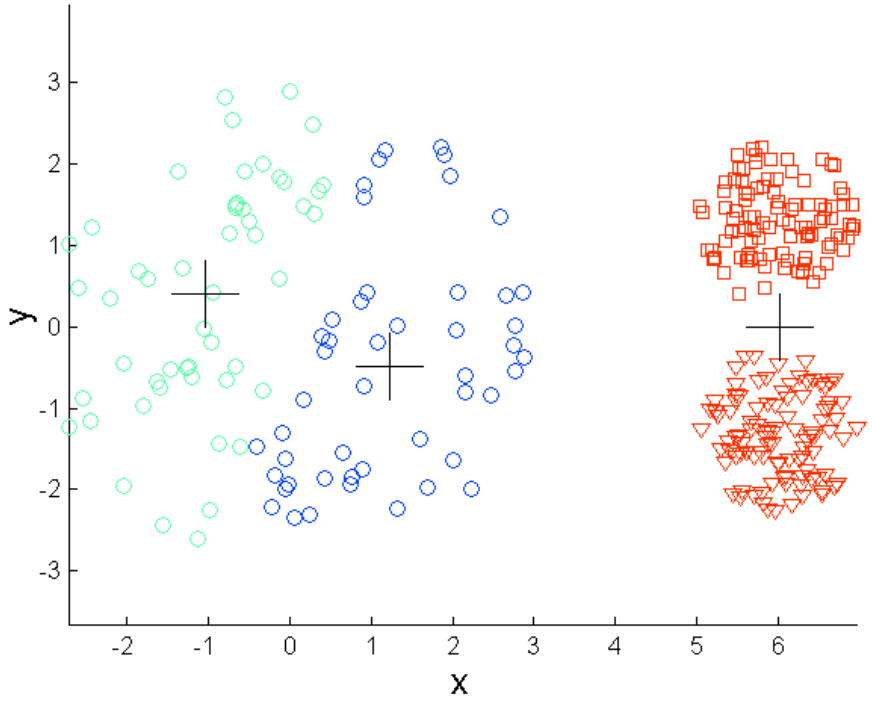
Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

K-Means Clustering: Weaknesses?

- Not robust when clusters have different densities:



Original Points

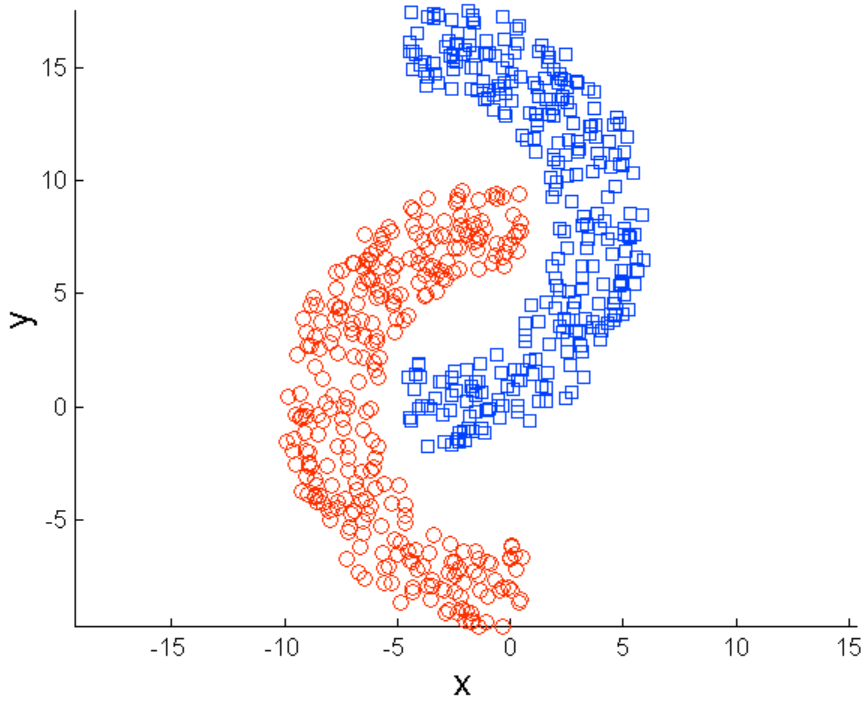


K-means (3 Clusters)

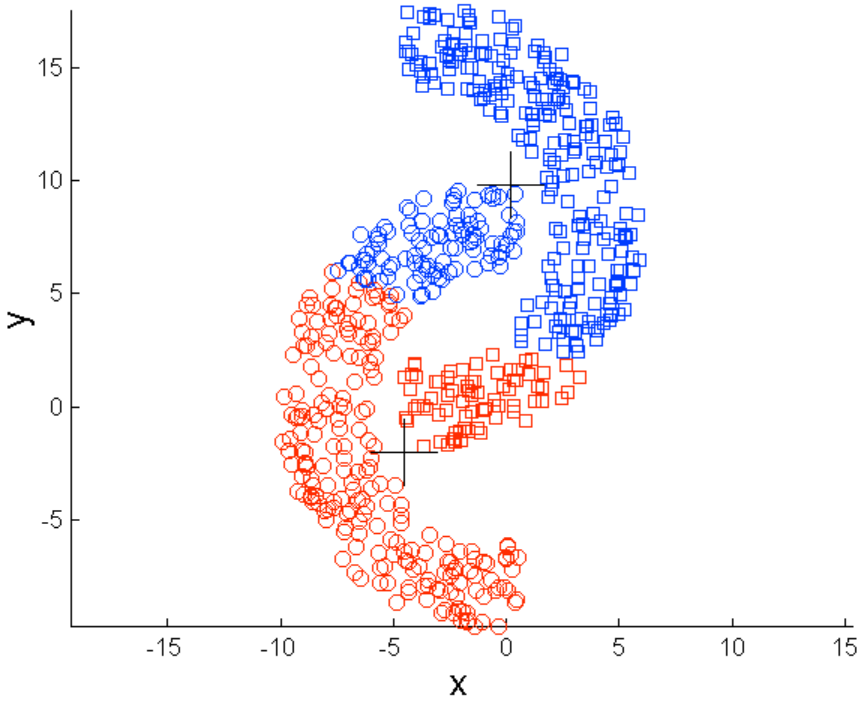
Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

K-Means Clustering: Weaknesses?

- Not robust when clusters have different globular shapes:



Original Points

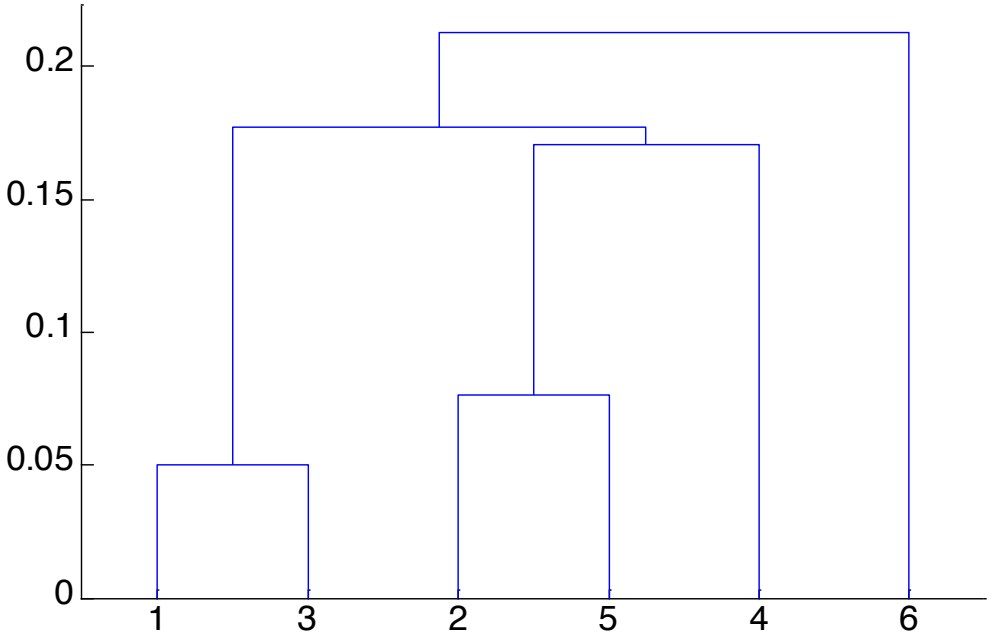


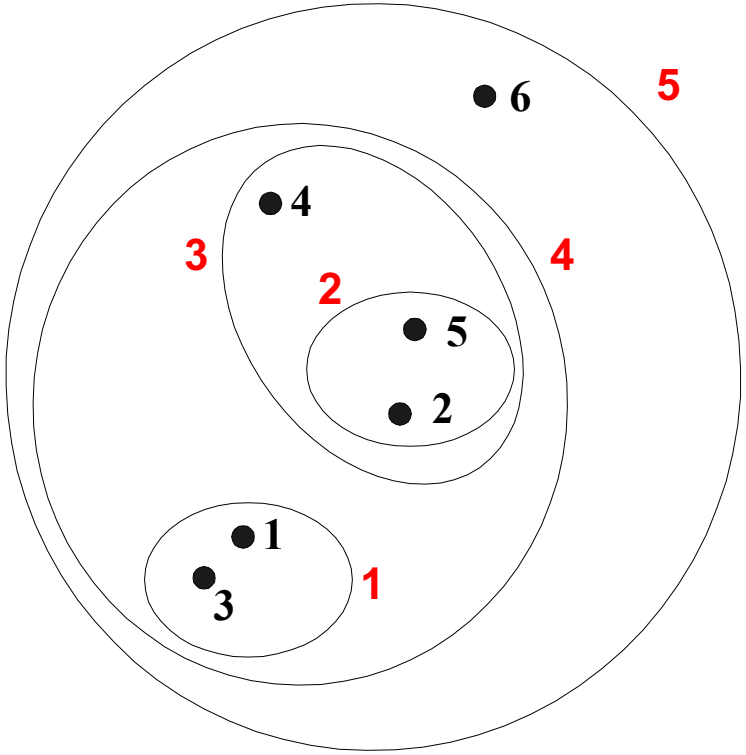
K-means (3 Clusters)

Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

Hierarchical Clustering

- Set of nested clusters organized in hierarchical tree by merging/splitting
- Dendrogram visualization: shows sequence of merges/splits





Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

Hierarchical Clustering: Two Main Approaches

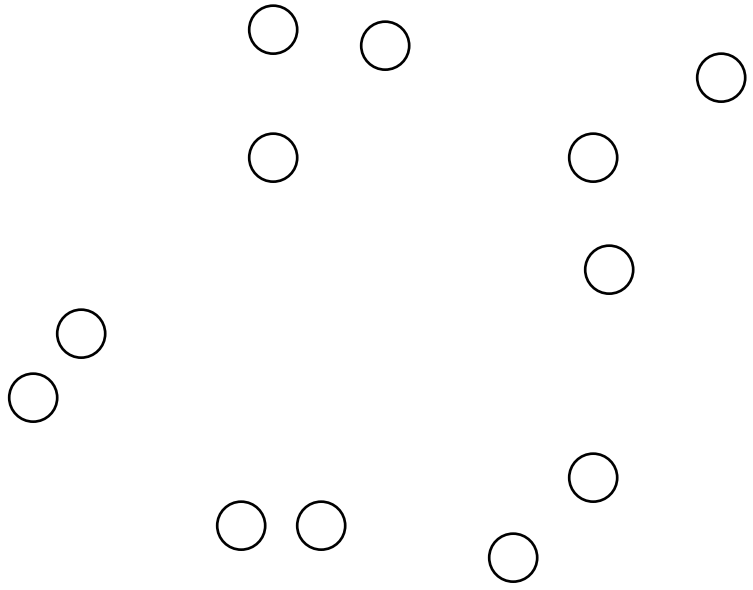
- Agglomerative:
  - Start with points as individual clusters
  - At each step, merge closest pair of clusters until only one cluster (or k clusters) left

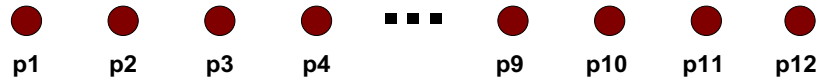
- Divisive:
  - Start with one, all-inclusive cluster
  - At each step, split a cluster until each cluster contains an individual point (or there are k clusters)

Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

Agglomerative Clustering: First Step

- Start with clusters of individual points and a proximity matrix

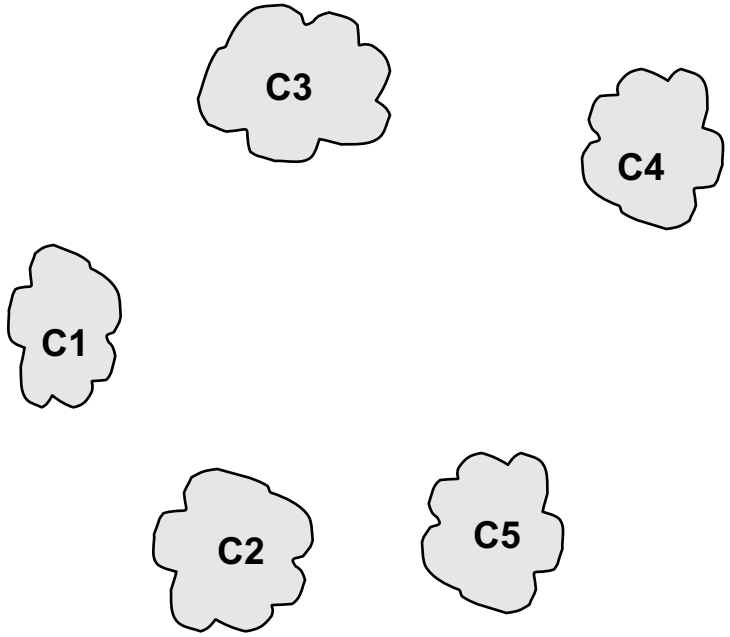


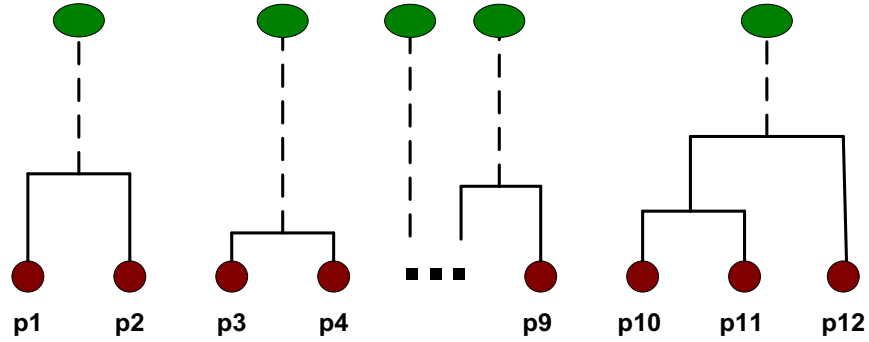


Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

Agglomerative Clustering: Intermediate Step

- Start with clusters of individual points

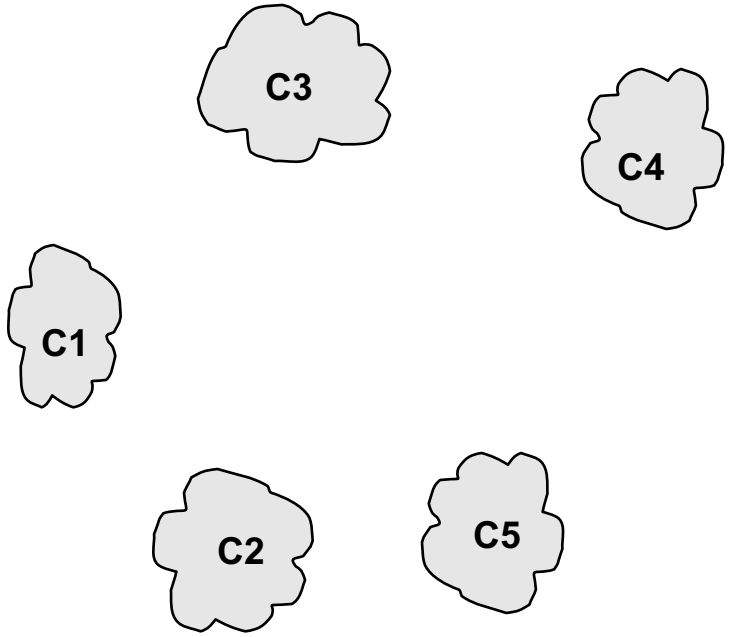


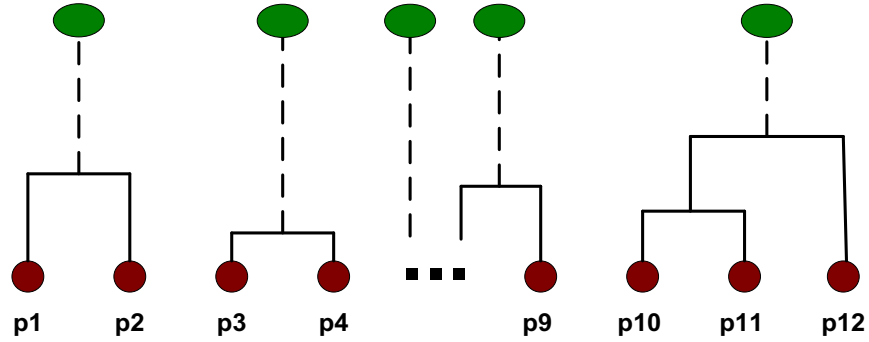


Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

Agglomerative Clustering: Intermediate Step

- After several merging steps, we have some clusters

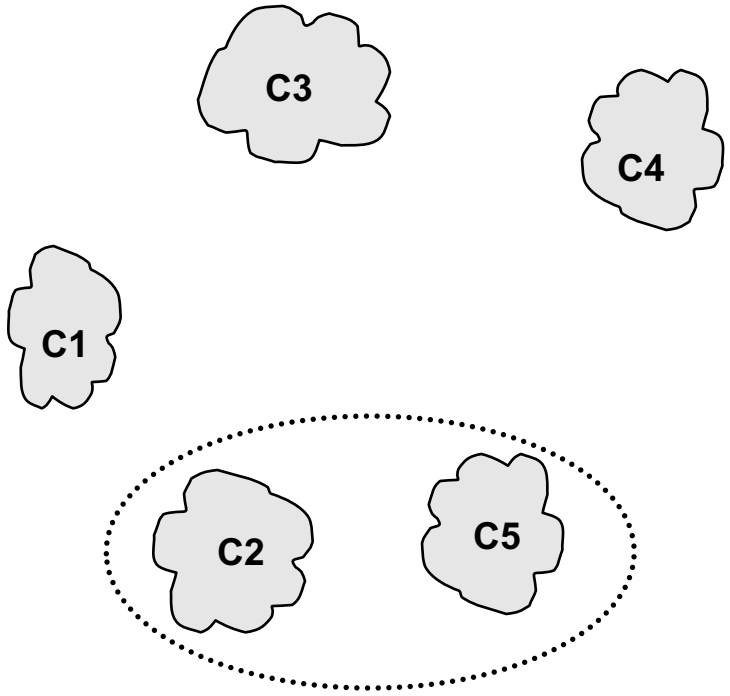


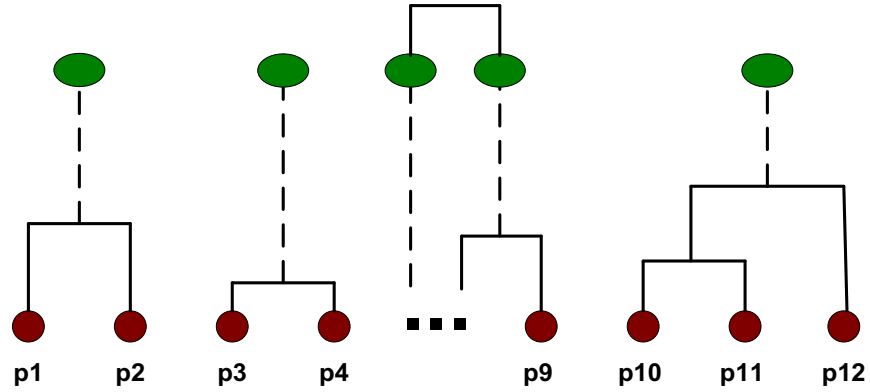


Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

Agglomerative Clustering: Intermediate Step

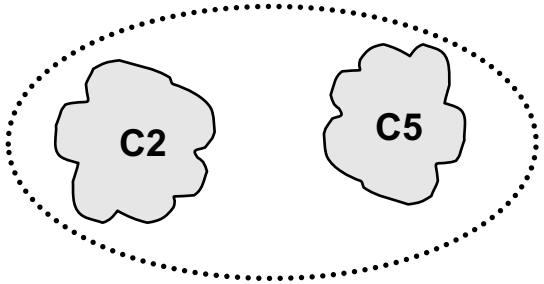
- Merge two closest clusters (C2 and C5)

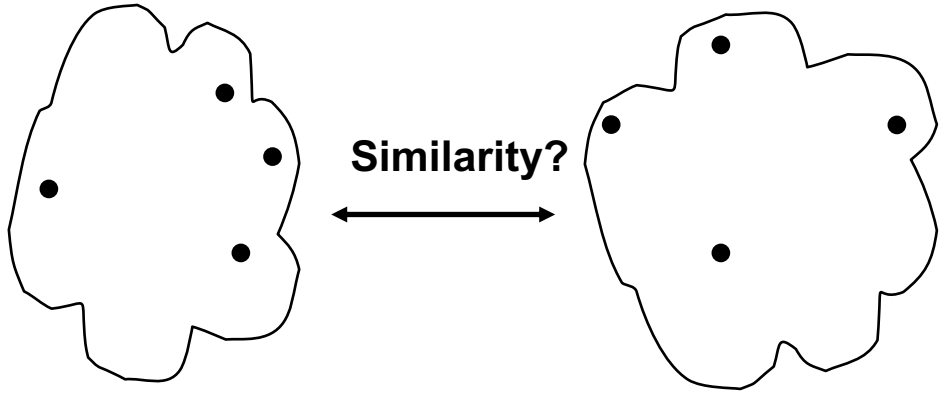




Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

How to Measure Inter-Cluster Distance?

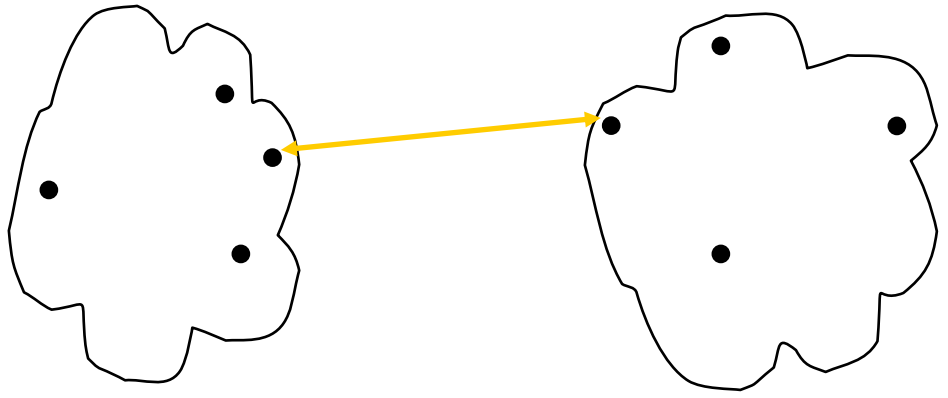




Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

How to Measure Inter-Cluster Distance?

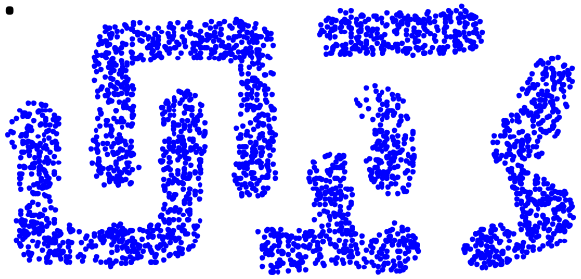
- Minimum distance



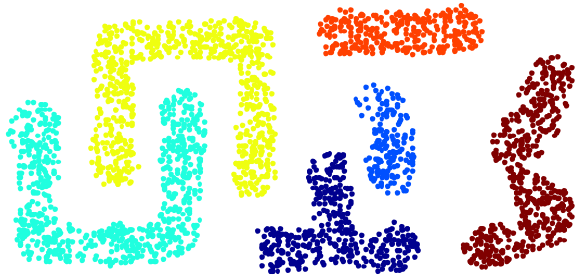
Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

Minimum Distance: Strengths/Weaknesses?

- Can handle non-elliptical shapes:

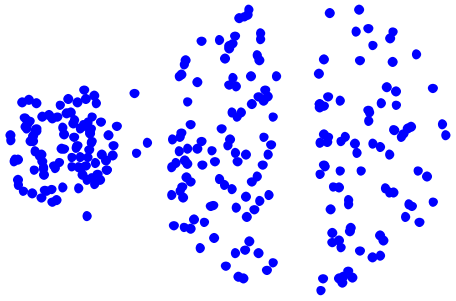


Original Points

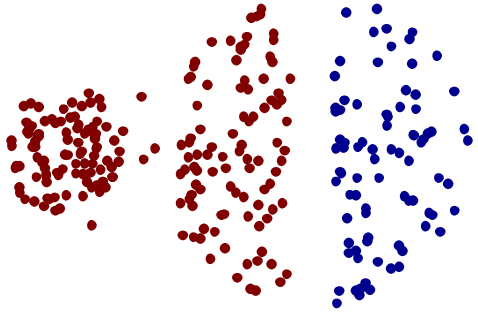


Six Clusters

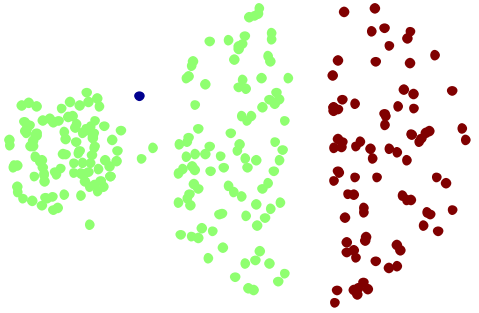
- Sensitive to noise and outliers:



Original Points



Two Clusters

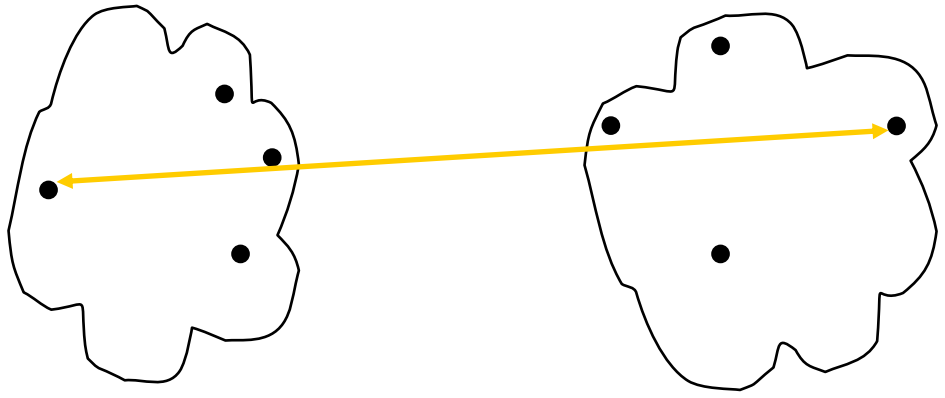


Three Clusters

Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

How to Measure Inter-Cluster Distance?

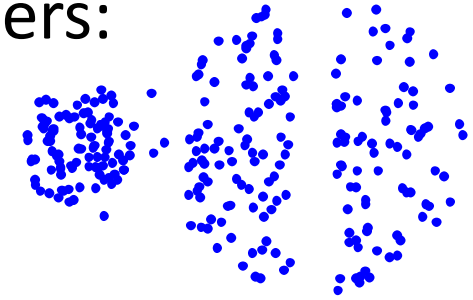
- Minimum distance
- Maximum distance



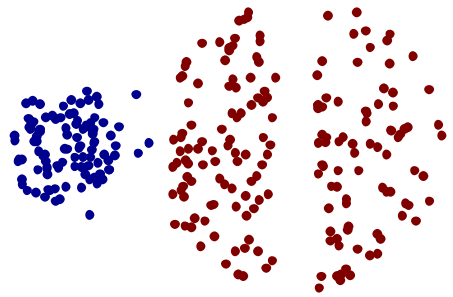
Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

Maximum Distance: Strengths/Weaknesses?

- Less susceptible to noise and outliers:

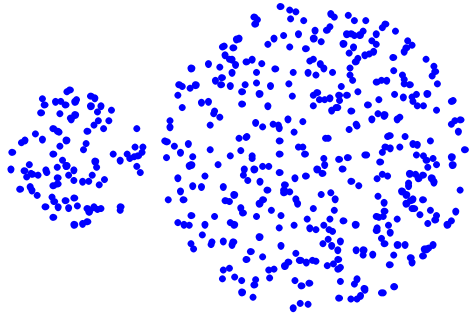


Original Points

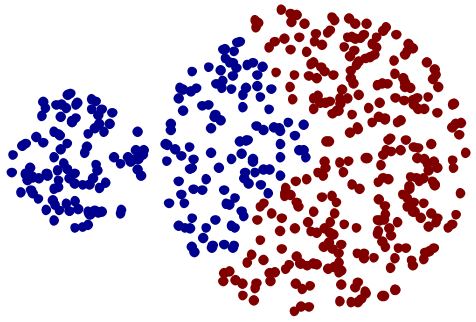


Two Clusters

- Tends to break large clusters:



Original Points

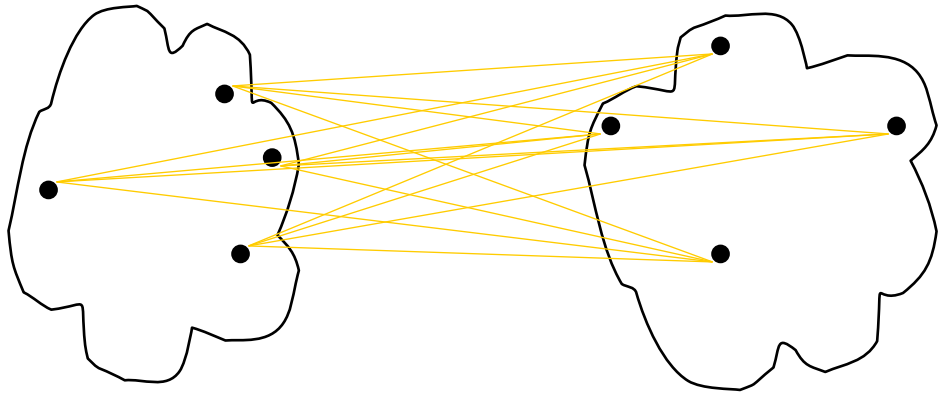


Two Clusters

Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

How to Measure Inter-Cluster Distance?

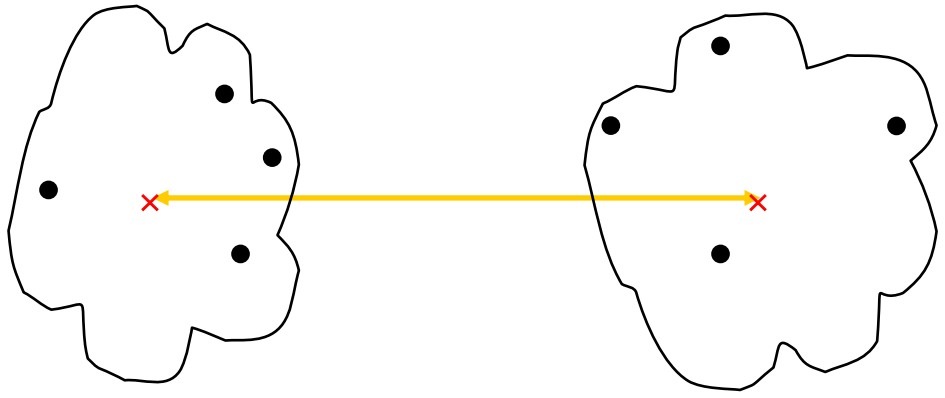
- Minimum distance
- Maximum distance
- Group average



Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

How to Measure Inter-Cluster Distance?

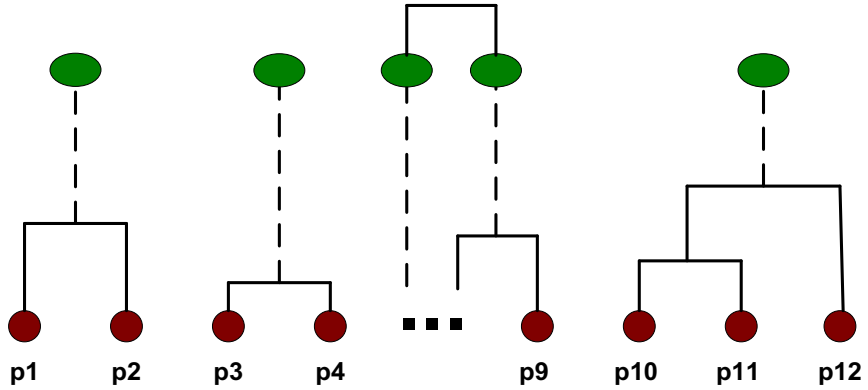
- Minimum distance
- Maximum distance
- Group average
- Distance Between Centroids



Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

Hierarchical Clustering: Strengths?

- Any number of clusters can be obtained by ‘cutting’ the dendrogram at the proper level



- They may correspond to meaningful taxonomies
  - Example in biological sciences (e.g., animal kingdom, phylogeny reconstruction, …)

Slide adapted from: https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap7_basic_cluster_analysis.pdf

Discussion

1. How is k-means clustering different from the K nearest neighbor algorithm?

2. After being disappointed with its service, Netflix hires you to build a new movie recommendations service. How would you design an unsupervised learning approach? And how would you evaluate your method?

Google Form: Guest Speaker & Class Feedback

- Google form
  - Guest: Dr. Suyog Jain, Senior Machine Learning Scientist at PathAI (http://suyogjain.com/): list one question for him for today's visit
  - Class feedback: Provide feedback about what, if anything, you would like to see improved with our remote learning course
- Then, take a short break.
- Class resumes at 4:50pm CST.

Today’s Topics

- Machine Learning for Unlabeled Data
- Autoencoders
- Clustering
- Guest: Dr. Suyog Jain from PathAI