CS540 Introduction to Artificial Intelligence

Lecture 15

Young Wu

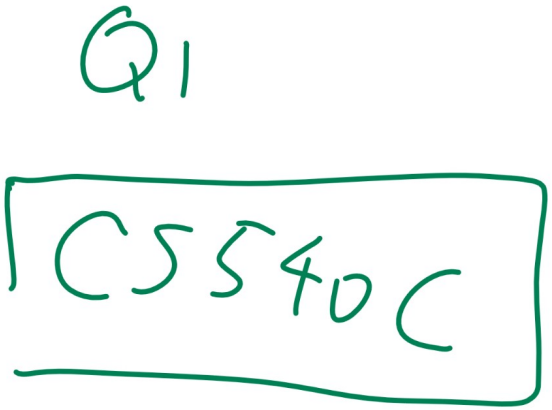
Based on lecture slides by Jerry Zhu, Yingyu Liang, and Charles Dyer

July 27, 2022

Midterm

Admin

- The midterm is:
- A : Too Easy
- B : Easy
- C : Just right
- D : Hard
- E : Too Hard



Midterm

Admin

- Stats for ID = test are posted, fix bugs in auto-grading, then post grades on Canvas.
- No curve for regular version midterm, may drop one question from make up midterm.
- Go over some midterm questions on Saturday, join if you have questions.
- Grades for $M1$ to $M7$, $P1$, $D1$ to $D3$ will be updated tomorrow night (last chance to submit these).
- Final exam is not cumulative, but some topics are still relavent (gradient descent, probability theory, reinforcement learning).

P2, P3

x/51 · 7.5

Unsupervised Learning

Motivation

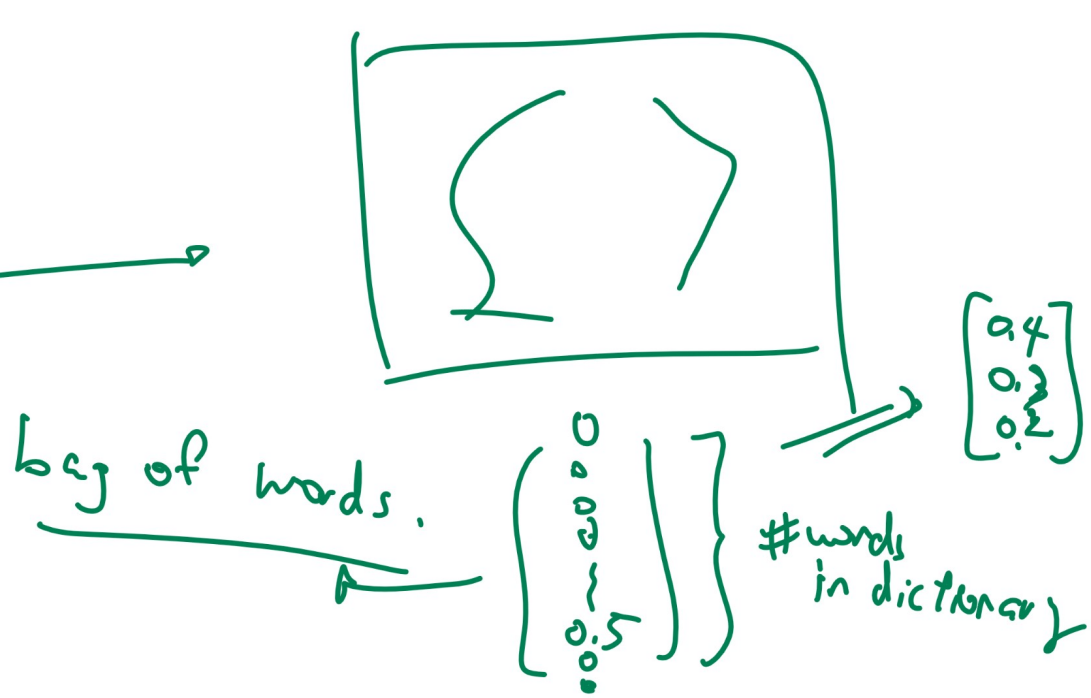
- Supervised learning: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
- Unsupervised learning: $x_1, x_2, \ldots, x_n$.
- There are a few common tasks without labels.

1. Clustering: separate instances into groups.
2. Novelty (outlier) detection: find instances that are different.
3. Dimensionality reduction: represent each instance with a lower dimensional feature vector while maintaining key characteristics.

Unsupervised Learning Applications

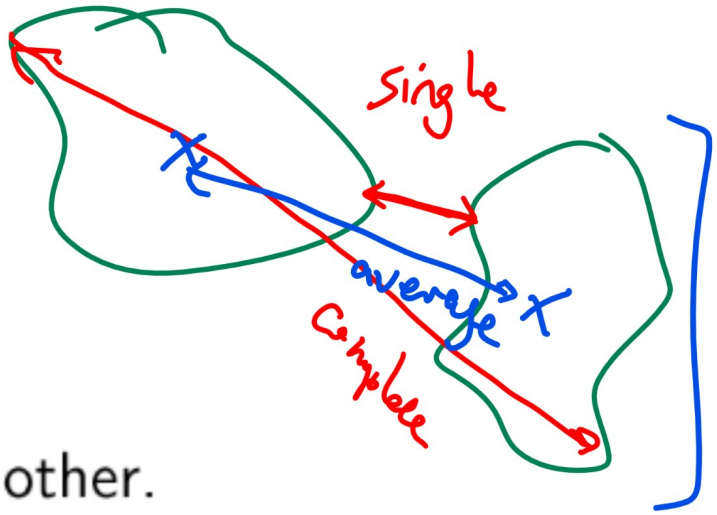
Motivation

1. Google News
2. Google Photo
3. Image Segmentation
4. Text Processing



Hierarchical Clustering

Description



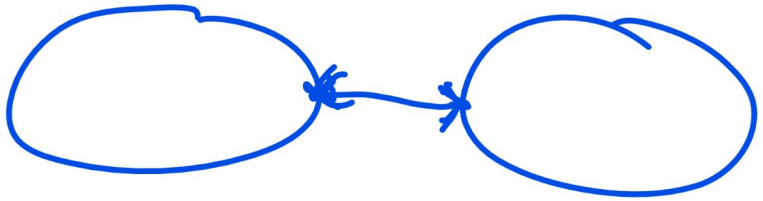
- Start with each instance as a cluster.
- Merge clusters that are closest to each other.
- Result in a binary tree with close clusters as children.

Hierarchical Clustering Diagram

Description

Single Linkage Distance

Definition



- Usually, the distance between two clusters is measured by the single-linkage distance.

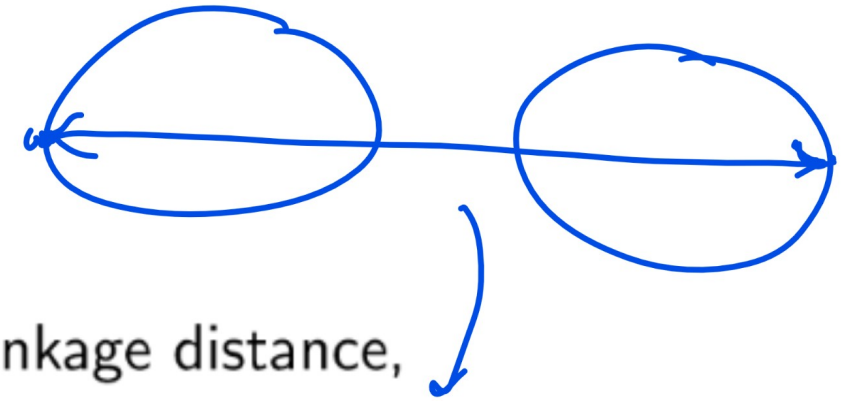
$$d(C_k, C_{k'}) = \min\{d(x_i, x_{i'}) : x_i \in C_k, x_{i'} \in C_{k'}\}$$

Manhattan, Euclidean. → P4

- It is the shortest distance from any instance in one cluster to any instance in the other cluster.

Complete Linkage Distance

Definition



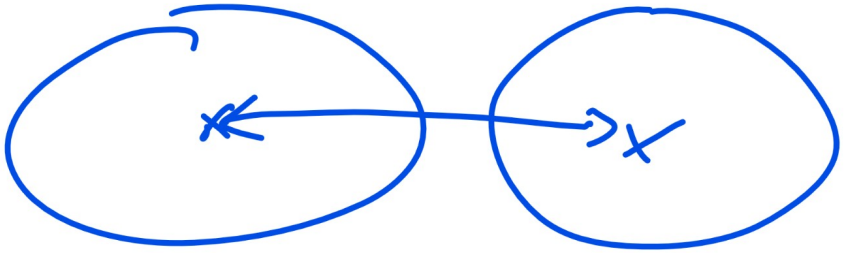
- Another measure is complete-linkage distance,

$$d(C_k, C_{k'}) = \max\{d(x_i, x_{i'}) : x_i \in C_k, x_{i'} \in C_{k'}\}$$

- It is the longest distance from any instance in one cluster to any instance in the other cluster.

Average Linkage Distance Diagram

Definition



- Another measure is average-linkage distance.

$$d(C_k, C_{k'}) = \frac{1}{|C_k||C_{k'}|} \sum_{x_i \in C_k, x_{i'} \in C_{k'}} d(x_i, x_{i'})$$

- It is the average distance from any instance in one cluster to any instance in the other cluster.

Hierarchical Clustering 1

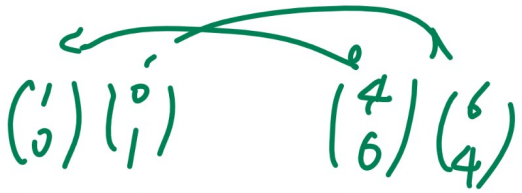
Quiz

- Given three clusters $A = \{0, 2, 6\}$, $B = \{3, 9\}$, $C = \{11\}$. What is the next iteration of hierarchical clustering with Euclidean distance and single and complete linkage?

	single	complete
→ AB	1	9
AC	5	11
BC	2	8
	merge AB	merge BC

Hierarchical Clustering 2

Quiz



Q2

- Given three clusters $A = \{0, 1\}, B = \{4, 6\}, C = \{8\}$. What is the next iteration of hierarchical clustering with Euclidean distance and complete linkage?
- A : Merge A and B.
- B : Merge A and C.
- C : Merge B and C.
- D : I don't understand.

AB 6

AC 8

BC 4

Hierarchical Clustering 3

Quiz

- Spring 2018 Midterm Q5
- Given three clusters $A = \{0, 1\}, B = \{4, 6\}, C = \{8\}$. What is the next iteration of hierarchical clustering with Euclidean distance and single linkage?
- A : Merge A and B.
- B : Merge A and C.
- C : Merge B and C.
- D : I don't understand.

Q3

AB 3

BC 2

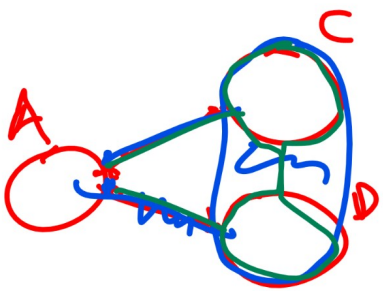
AC 7

Hierarchical Clustering 4

Quiz

- Given the distance between the clusters so far. Which pair of clusters will be merged using single linkage.

–	A	B	C	D	E
A	0	1075	2013	2054	996
B	1075	0	3272	2687	2037
C	2013	3272	0	808	1307
D	2054	2687	808	0	1059
E	996	2037	1307	1059	0



merge CD

A B CD E

1075 2013 996

2687 2037

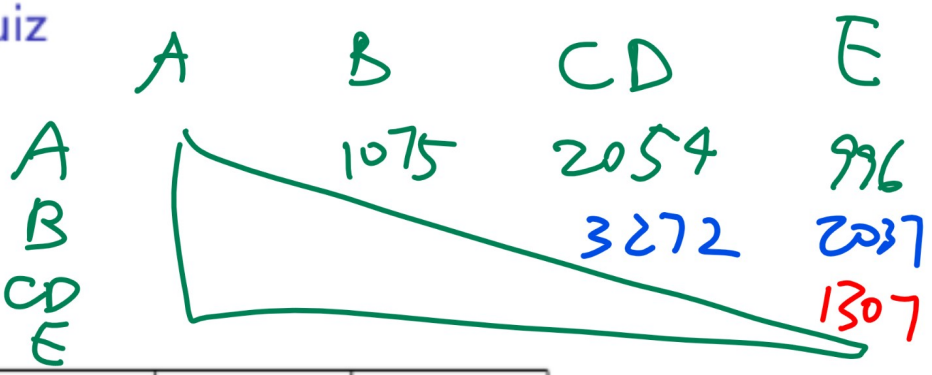
1059

A B CDE

Hierarchical Clustering 4, Diagram

Quiz

Comple



–	A	B	C	D	E
A	0	1075	2013	2054	996
B	1075	0	3272	2687	2037
C	2013	3272	0	808	1307
D	2054	2687	808	0	1059
E	996	2037	1307	1059	0

P4

M8

Hierarchical Clustering 5

Quiz

- Given the distance between the clusters so far. Which pair of clusters will be merged using complete linkage.

–	A	*B*	*C*	*D*
A	0	1075	2013	2054
B	1075	0	3272	2687
C	2013	3272	0	808
D	2054	2687	808	0

MJQ.

- *E* : *I* don't understand.

Number of Clusters

Discussion

- K can be chosen using prior knowledge about X.
- The algorithm can stop merging as soon as all the between-cluster distances are larger than some fixed R.
- The binary tree generated in the process is often called dendrogram, or taxonomy, or a hierarchy of data points.
- An example of a dendrogram is the tree of life in biology.

not pick K.

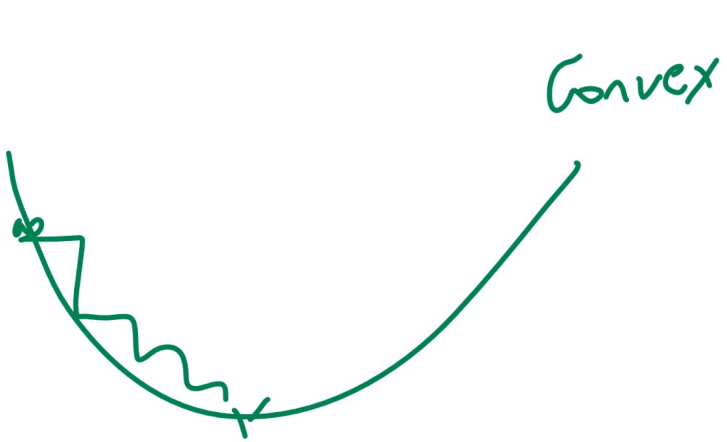
K Means Clustering

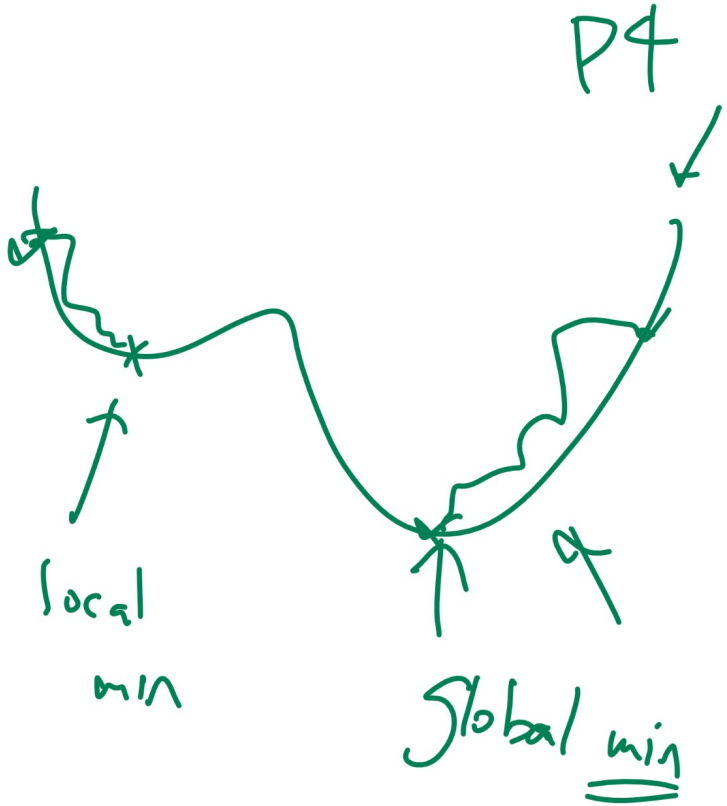
Description

- This is not *K* Nearest Neighbor.
- Start with random cluster centers.
- Assign each point to its closest center.
- Update all cluster centers as the center of its points.

K Means Clustering Demo

Description





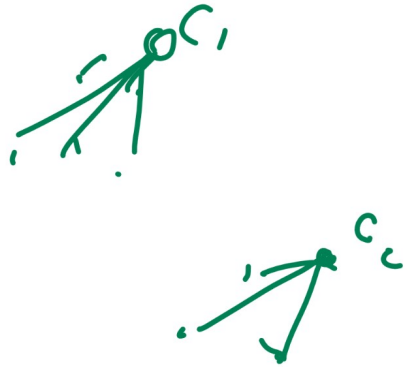
Distortion

Distortion

- Distortion for a point is the distance from the point to its cluster center.
- Total distortion is the sum of distortion for all points.

$$D_K = \sum_{i=1}^{n} d\left(x_i, c_{k^\star(x_i)}(x_i)\right)$$

$$k^\star(x) = \underset{k=1,2,\ldots K}{\operatorname{argmin}}\, d(x, c_k)$$



Objective Function Counterexample

Definition

Gradient Descent

Definition

- When d is the Euclidean distance. K Means algorithm is the gradient descent when distortion is the objective (cost) function.

$$\frac{\partial}{\partial c_k} \sum_{k=1}^{K} \sum_{x \in C_k} \|x - c_k\|_2^2 = 0$$

min total distortion

$$\Rightarrow -2 \sum_{x \in C_k} (x - c_k) = 0$$

$$\Rightarrow c_k = \frac{1}{|C_k|} \sum_{x \in C_k} x$$

K-means

K Means Clustering 1

Quiz

- Given data $x = \{-1, 0, 2\}$ and initial cluster centers $c_1 = 0, c_2 = 1$, what is the initial clusters and what is the initial total distortion (sum of squares without square root)?

	c_1	c_2	which
-1	1	2	c_1
0	0	1	c_1
2	2	1	c_2

$\{-1, 0\}$. $\{2\}$

$c_1' = -\frac{1}{2}$ $c_2' = 2$

$1^2 + 0^2 + 1^2 = 2$

$(\frac{1}{2})^2 + (\frac{1}{2})^2 + 0^2 = \frac{1}{2}$

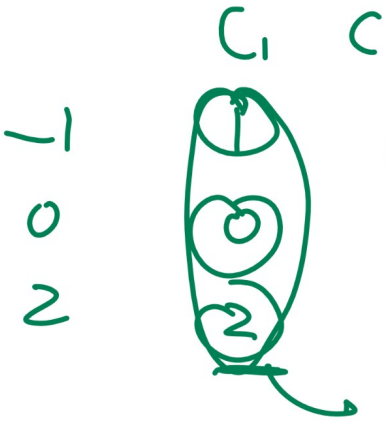
K Means Clustering 2

Quiz

Q4

- Given data $x = \{-1, 0, 2\}$ and initial cluster centers $c_1 = 0, c_2 = 5$, what is the initial clusters?
- $A : \{\emptyset\}$ and $\{-1, 0, 2\}$
- $B : \{-1\}$ and $\{0, 2\}$
- $C : \{-1, 0\}$ and $\{2\}$
- $D : \{-1, 0, 2\}$ and $\{\emptyset\}$
- E : I don't understand.

total distortion



Total Distortion 2

Quiz

Q5

- Given data $x = \{-1, 0, 2\}$ and initial cluster centers $c_1 = 0, c_2 = 5$, what is the initial total distortion (sum of squares without square root)?
- $A : 2$
- $B : 5$
- $C : 10$
- $D : 50$
- $E : I$ don't understand.

$1^2 + 0^2 + 2^2 = 5$

Number of Clusters

Discussion

- There are a few ways to pick the number of clusters K.

1. K can be chosen using prior knowledge about X.
2. K can be the one that minimizes distortion? No, when $K = n$, distortion $= 0$.
3. K can be the one that minimizes distortion + regularizer.

$$K^\star = \underset{k}{\operatorname{argmin}} \left(D_k + \lambda \cdot m \cdot k \cdot \log n \right)$$

features

items

regularization param.

- λ is a fixed constant chosen arbitrarily.

Initial Clusters

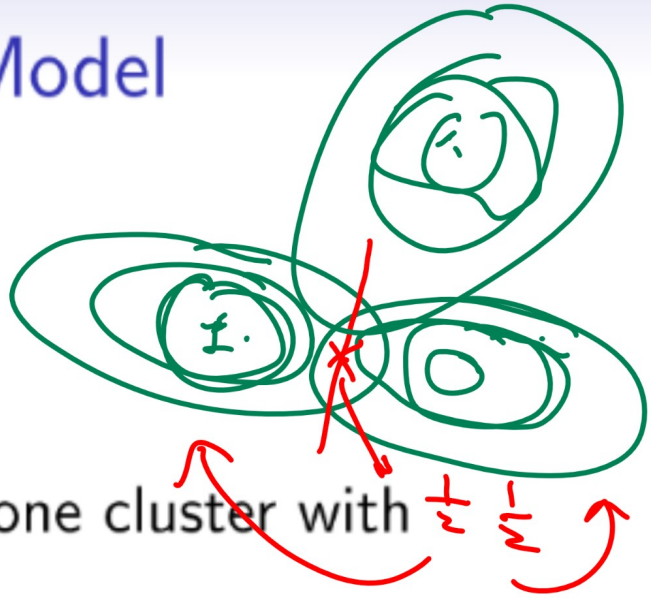
Discussion

- There are a few ways to initialize the clusters.

1. K uniform random points in $\{x_i\}_{i=1}^n$.
2. 1 uniform random point in $\{x_i\}_{i=1}^n$ as $c_1^{(0)}$, then find the farthest point in $\{x_i\}_{i=1}^n$ from $c_1^{(0)}$ as $c_2^{(0)}$, and find the farthest point in $\{x_i\}_{i=1}^n$ from the closer of $c_1^{(0)}$ and $c_2^{(0)}$ as $c_3^{(0)}$, and repeat this K times.

Gaussian Mixture Model

Discussion



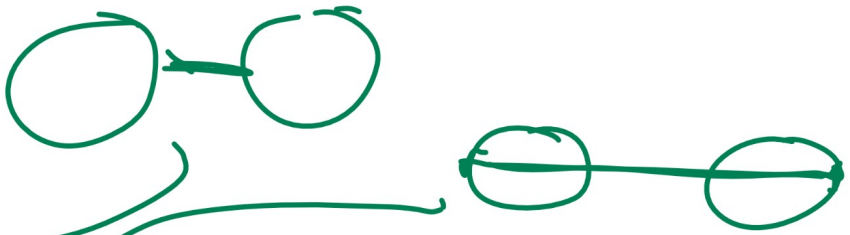
- In K means, each instance belong to one cluster with certainty.
- One continuous version is called the Gaussian mixture model: each instance belongs to one of the clusters with a positive probability.
- The model can be trained using Expectation Maximization Algorithm (EM Algorithm).

Gaussian Mixture Model Demo

Discussion

Summary

Description



- Unsupervised learning:
  1. Clustering: Hierachical → Start with singleton clusters → Merge closest (single, complete linkage) clusters → Repeat.
  2. Clustering: K-Means → Start with random centers → Find closest center to every point → Update centers → Repeat.
  3. Dimensionality Reduction: Principal Component Analysis.