Unsupervised Learning
ooo

Hierarchical Clustering
ooooooooooooooo

K Means Clustering
ooooooooooooooo

CS540 Introduction to Artificial Intelligence
Lecture 11

Young Wu

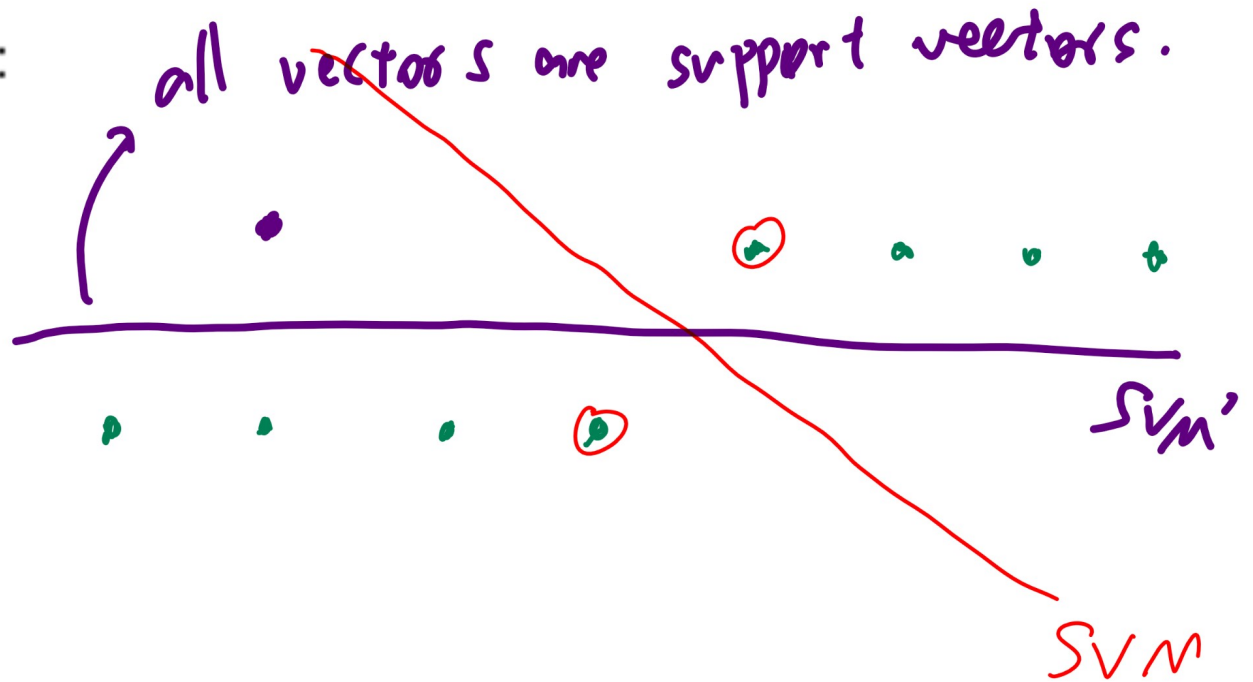Based on lecture slides by Jerry Zhu, Yingyu Liang, and Charles Dyer

July 13, 2020

Unsupervised Learning
●○○

Hierarchical Clustering
○○○○○○○○○○○○○○

K Means Clustering
○○○○○○○○○○○○○○○

Midterm
Admin

Socrative
Room CS540E

Q1

- The midterms are:
- A: Too Easy
- B: Easy
- C: $\updownarrow$
- D: Hard
- E: Too Hard

all vectors are support vectors.

SVM'

SVM

Unsupervised Learning
○●○

Hierarchical Clustering
○○○○○○○○○○○○○○

K Means Clustering
○○○○○○○○○○○○○○○

Midterm Discussion
Admin

- Bug fix in auto-grading, grades updated.
- Did not fix individual grades.
- Version A Part 1 average: 4.5, Part 2 average: 3.8
- Version B Part 1 average: 4.1, Part 2 average: 3.7
- None of the questions has PROB < 0.25, RPBI < 0.
- No curve for all versions.

MIB

cross entropy

Unsupervised Learning
○○●

Hierarchical Clustering
○○○○○○○○○○○○○○

K Means Clustering
○○○○○○○○○○○○○○

Unsupervised Learning
Motivation

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

1. Clustering: separate instances into groups.

   $0, 1, 2, 3,$

2. Novelty (outlier) detection: find instances that are different.

   $0, 1$

3. Dimensionality reduction: represent each instance with a lower dimensional feature vector while maintaining key characteristics.
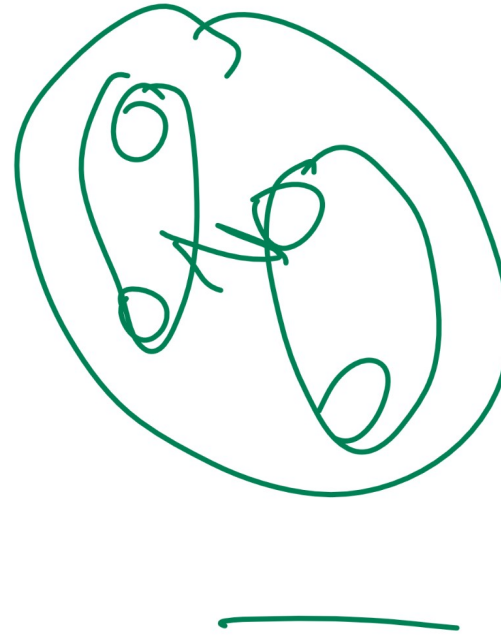
   $0, 0.1, 0.2$.

Unsupervised Learning
○○○

Hierarchical Clustering
●○○○○○○○○○○○○○

K Means Clustering
○○○○○○○○○○○○○○○

Hierarchical Clustering

Description

- Start with each instance as a cluster.

- Merge clusters that are closest to each other.

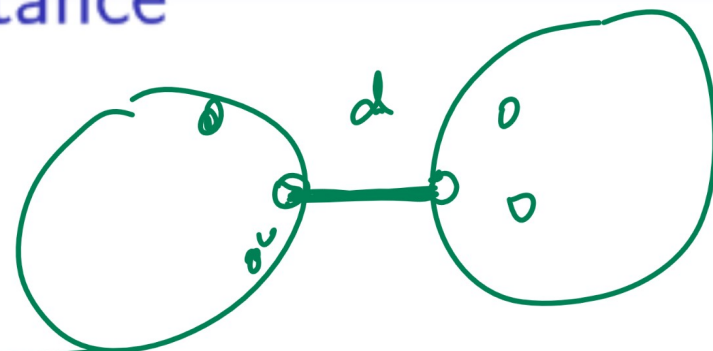- Result in a binary tree with close clusters as children.

Unsupervised Learning
○○○

Hierarchical Clustering
○●○○○○○○○○○○○○

K Means Clustering
○○○○○○○○○○○○○○○

Hierarchical Clustering Diagram

Description

Unsupervised Learning
ooo

Hierarchical Clustering
oo●oooooooooooo

K Means Clustering
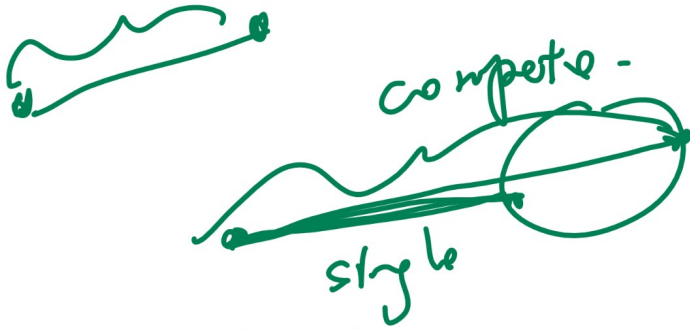ooooooooooooooo

Single Linkage Distance

Definition

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

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

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

Unsupervised Learning
ooo

Hierarchical Clustering
ooo●ooooooooooo

K Means Clustering
ooooooooooooooo

Complete Linkage Distance

Definition

- Another measure is complete-linkage distance,

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

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

Unsupervised Learning
○○○

Hierarchical Clustering
○○○○●○○○○○○○○○

K Means Clustering
○○○○○○○○○○○○○○○○

Average Linkage Distance Diagram
Definition

- Another measure is average-linkage distance.

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

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

not used.

Unsupervised Learning
○○○

Hierarchical Clustering
○○○○○●○○○○○○

K Means Clustering
○○○○○○○○○○○○○

Hierarchical Clustering 1

Quiz

always merge closest Clusters
with d

- Spring 2018 Midterm Q5

- 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 linkage?

single   complete
  d d

- A: Merge A and B.

- B: Merge A and C.

- C: Merge B and C.

	single	complete
AB	1	9
AC	5	11
BC	2	8

if complete linkage merge B, C.

Unsupervised Learning
ooo

Hierarchical Clustering
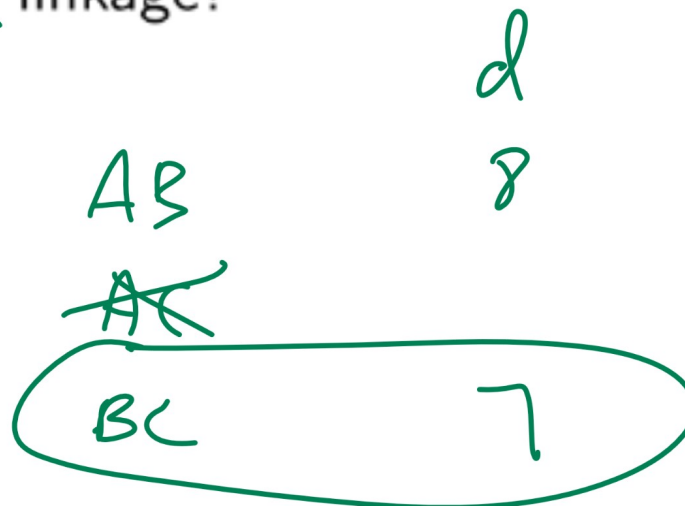oooooo●ooooooo

K Means Clustering
ooooooooooooooo

Hierarchical Clustering 2

Quiz

Q2

- Spring 2018 Midterm Q5
- Given three clusters $A = \{0, 1\}$, $B = \{4, 8\}$, $C = \{10, 11\}$. 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

AB   *8*

AC

BC   *7*

Unsupervised Learning
ooo

Hierarchical Clustering
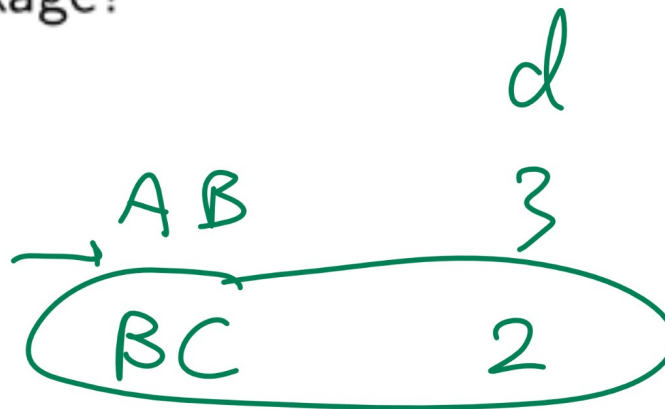ooooooo●oooooo

K Means Clustering
oooooooooooooo

Hierarchical Clustering 3

Quiz

Q3

- Spring 2018 Midterm Q5
- Given three clusters $A = \{0, 1\}$, $B = \{4, 8\}$, $C = \{10, 11\}$. 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

A B   3

BC   2

tie break by combing with lower indexed cluster

Unsupervised Learning

○○○

Hierarchical Clustering

○○○○○○○○●○○○○○

K Means Clustering

○○○○○○○○○○○○○○
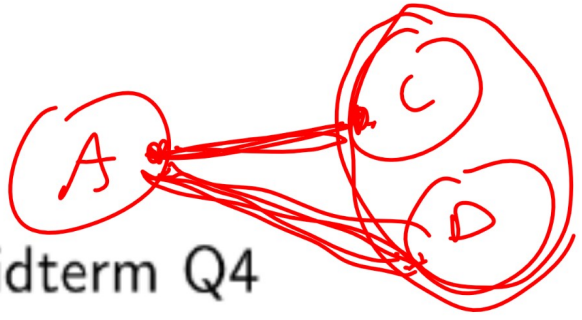
Hierarchical Clustering 4

Quiz

- Spring 2018 Midterm Q5
- Given three clusters $A = \{0, 2, 6\}$, $B = \{3, 9\}$, $C = \{11\}$. 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.

Unsupervised Learning
ooo

Hierarchical Clustering
ooooooooo●oooo

K Means Clustering
ooooooooooooooo

Hierarchical Clustering 3

Quiz

A   B   C-D   E

A   0   1075   2013   996
      0   2687   2037
          0   1059
              0

P4

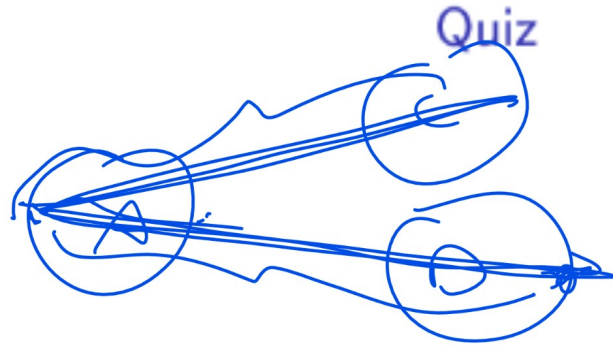- Spring 2017 Midterm Q4

- 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

C-D

merge   C and D

Unsupervised Learning
ooo

Hierarchical Clustering
oooooooooooo●ooo

K Means Clustering
oooooooooooooooo

Hierarchical Clustering 4

Quiz

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

Unsupervised Learning
○○○

Hierarchical Clustering
○○○○○○○○○○○○●○○

K Means Clustering
○○○○○○○○○○○○○○○

Hierarchical Clustering 5
Quiz

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

−	A	B	C	E
A	0	1075	2013	996
B	1075	0	2687	2037
C	2013	2687	0	1059
E	996	2037	1059	0

Unsupervised Learning
ooo

Hierarchical Clustering
ooooooooooooo●o

K Means Clustering
ooooooooooooooo

Hierarchical Clustering

Algorithm

- Input: instances: $\{x_i\}_{i=1}^n$, the number of clusters K, and a distance function d.

- Output: a list of clusters $C = C_1, C_2, ..., C_K$.

- Initialize for $t = 0$.

$$C^{(0)} = C_1^{(0)}, ..., C_n^{(0)}, \text{ where } C_k^{(0)} = \{x_k\}, k = 1, 2, ..., n$$

- Loop for $t = 1, 2, ..., n - k + 1$.

$$(k_1^\star, k_2^\star) = \arg\min_{k_1, k_2} d\left(C_{k_1}^{(t-1)}, C_{k_2}^{(t-1)}\right)$$

$$C^{(t)} = \left(C_{k_1^\star}^{(t-1)} \cup C_{k_2^\star}^{(t-1)}\right), C_1^{(t-1)}, ... \text{ no } k_1^\star, k_2^\star ..., C_n^{(t-1)}$$

Unsupervised Learning
○○○

Hierarchical Clustering
○○○○○○○○○○○○○●

K Means Clustering
○○○○○○○○○○○○○○○

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.

Unsupervised Learning
○○○

Hierarchical Clustering
○○○○○○○○○○○○○○

K Means Clustering
●○○○○○○○○○○○○○○

K Means Clustering

Description

back

6:25

- 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.

Unsupervised Learning
○○○

Hierarchical Clustering
○○○○○○○○○○○○○○

K Means Clustering
○●○○○○○○○○○○○○○○

K Means Clustering Diagram

Description

Unsupervised Learning
○○○

Hierarchical Clustering
○○○○○○○○○○○○○○

K Means Clustering
○○●○○○○○○○○○○○○

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.

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

$$k^*(x) = \arg\min_{k=1,2,\dots K} d(x, c_k)$$

Unsupervised Learning
000

Hierarchical Clustering
00000000000000

K Means Clustering
000●0000000000

Objective Function

Definition

- When using Euclidean distance, sometimes total distortion is defined as sum of squared distances.

$$D_K = \sum_{i=1}^{n} d_2 \left(x_i, c_{k^\star(x_i)}(x_i)\right)^2$$

- This algorithm stop in finite steps.
- This algorithm is trying to minimize the total distortion but fails.

local min.

Unsupervised Learning
○○○

Hierarchical Clustering
○○○○○○○○○○○○○○

K Means Clustering
○○○○○●○○○○○○○○○

Objective Function Counterexample
Definition

Unsupervised Learning
ooo

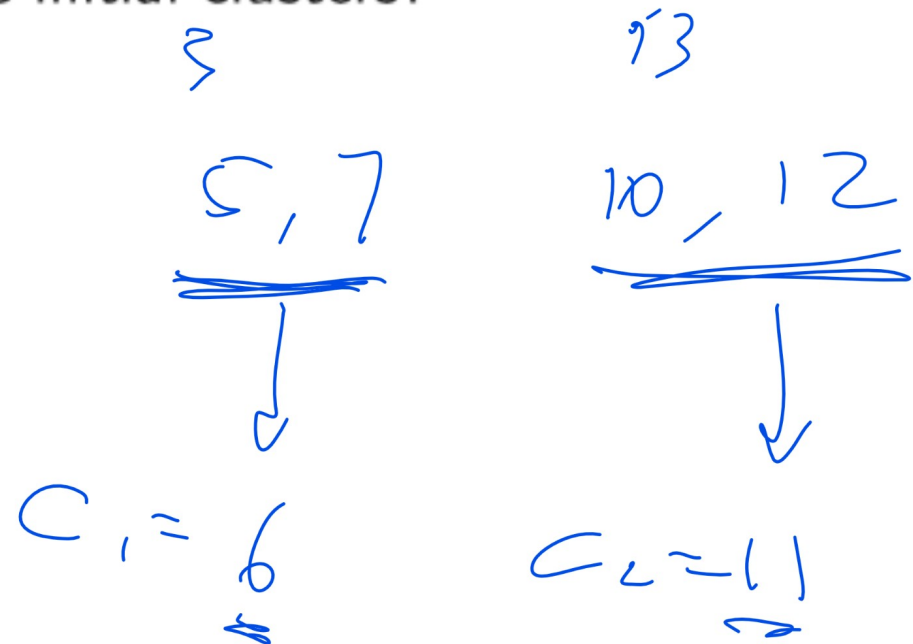Hierarchical Clustering
oooooooooooooo

K Means Clustering
ooooo●ooooooo

K Means Clustering 1

Quiz

- Spring 2018 Midterm Q5
- Given data $\{5, 7, 10, 12\}$ and initial cluster centers $c_1 = 3, c_2 = 13$, what is the initial clusters?
- A: $\{5, 7\}$ and $\{10, 12\}$
- B: $\{5\}$ and $\{7, 10, 12\}$
- C: $\{5, 7, 10\}$ and $\{12\}$

3 13

$5, 7$ $10, 12$

$C_1 = 6$ $C_2 = 11$

Unsupervised Learning
ooo

Hierarchical Clustering
ooooooooooooooo

K Means Clustering
oooooo●ooooooo

K Means Clustering 2

Quiz

$$\text{initial distortion} = 2 + 4 + 3 + 1$$

- Spring 2018 Midterm Q5
- Given data $\{5, 7, 10, 12\}$ and initial cluster centers $c_1 = 3, c_2 = 13$, what are the clusters in the next iteration?
- A: $\{5, 7\}$ and $\{10, 12\}$
- B: $\{5\}$ and $\{7, 10, 12\}$
- C: $\{5, 7, 10\}$ and $\{12\}$

$$5 \quad 7 \qquad 10, 12$$

$$c_1 = 6 \qquad\qquad c_2 = 11$$

$$\text{total distortion} \quad 1 + 1 \quad + \quad 1 + 1$$

$$= 4$$

Unsupervised Learning
○○○

Hierarchical Clustering
○○○○○○○○○○○○○○

K Means Clustering
○○○○○○○●○○○○○○

K Means Clustering 3

Quiz

Q4

- Given data -2, 0, 10 and initial cluster centers $c_1 = -4$, $c_2 = 1$, what is the initial clusters?

- A: $\{\varnothing\}$ and -2, 0, 10

- B: -2 and $\{0, 10\}$

- C: -2, 0 and $\{10\}$

- D: -2, 0, 10 and $\{\varnothing\}$

4

Unsupervised Learning
ooo

Hierarchical Clustering
oooooooooooooo

K Means Clustering
ooooooooo●oooo

Total Distortion 1

Quiz

Q5

$2 + 1 + 9$

- Given data $-2, 0, 10$ and initial cluster centers $c_1 = -4, c_2 = 1$, what is the initial total distortion?

Manhattan distance,

- A: 0

- B: 2

- ✓ C: 12

- D: 13

- E: 15

Unsupervised Learning
ooo   ooo

Hierarchical Clustering
ooooooooooooooo

K Means Clustering
oooooooooo●oooo
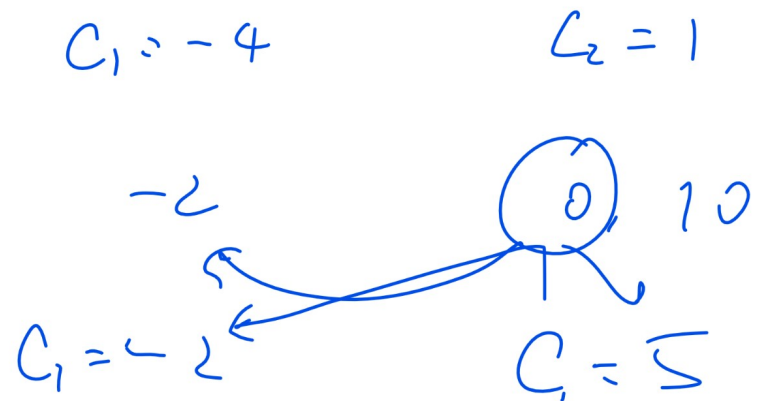
K Means Clustering 4

Quiz

Q6

- Given data $-2, 0, 10$ and initial cluster centers $c_1 = -4, c_2 = 1$, what are the clusters in the next iteration?

- A: $\{\varnothing\}$ and $-2, 0, 10$
- B: -2 and $\{0, 10\}$
- C: $-2, 0$ and $\{10\}$
- D: $-2, 0, 10$ and $\{\varnothing\}$

$c_1 = -4$ $\qquad$ $c_2 = 1$

-2 $\qquad$ 0 $\quad$ 10

$c_1 = -2$ $\qquad$ $c_2 = 5$

$|-2 - (-11)| + |0 - (-1)|$
$+ |10 - 10|$ (2) 2 to

-2 $\quad 0$

$c_1 = -1$ $\qquad$ 10 $\quad$ $c_2 = 10$

Unsupervised Learning
ooo●

Hierarchical Clustering
oooooooooooooo

K Means Clustering
ooooooooooo●ooo

Total Distortion 2

Quiz

- Given data -2, 0, 10 and initial cluster centers $c_1 = -4, c_2 = 1$, what is the final total distortion?
- A: 0
- B: 2
- C: 12
- D: 13
- E: 15

K Means Clustering

Algorithm

- Input: instances: $\{x_i\}_{i=1}^n$, the number of clusters K, and a distance function d.

- Output: a list of clusters $C = C_1, C_2, ..., C_K$.

- Initialize $t = 0$.

$$c_k^{(0)} = K \text{ random points}$$

- Loop until $c^{(t)} = c^{(t-1)}$.

$$C_k^{(t-1)} = \left\{ x : k = \arg\min_{k' \in 1,2,...,K} d\left(x, c_k^{(t-1)}\right) \right\}$$

$$c_k^{(t)} = \frac{1}{\left|C_k^{(t-1)}\right|} \sum_{x \in C_k^{(t-1)}} x$$

Unsupervised Learning
○○○

Hierarchical Clustering
○○○○○○○○○○○○○○○

K Means Clustering
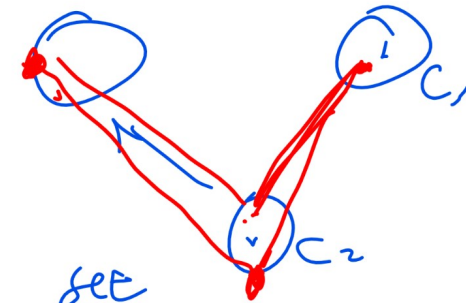○○○○○○○○○○○○○●○

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.

$$K^\star = \arg\min_k \left(D_k + \lambda \cdot m \cdot k \cdot \log n \right)$$

- λ is a fixed constant chosen arbitrarily.

Initial Clusters

Discussion

handwritten: at 7:00

handwritten: P4

handwritten: in their ming set

- 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.

handwritten: C_1 C_2 C_1