UCLA

CS145 Discussion

Week 6

Junheng, Shengming, Yunsheng

11/09/2018

UCLA

Today's Discussion

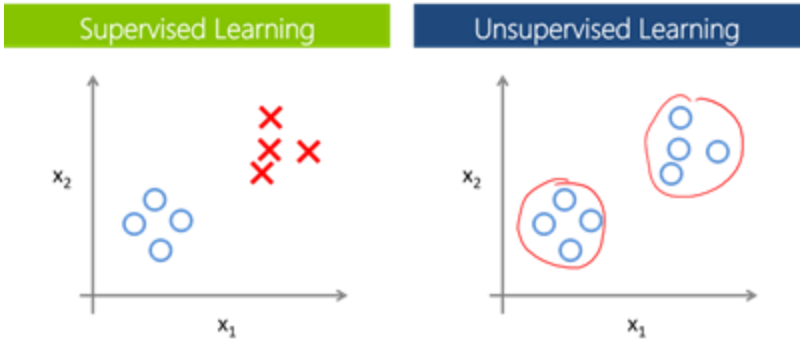
- Announcement
- Roadmap
  - Clustering: K-means, DB-SCAN
  - Midterm Review
- Q & A

Announcement

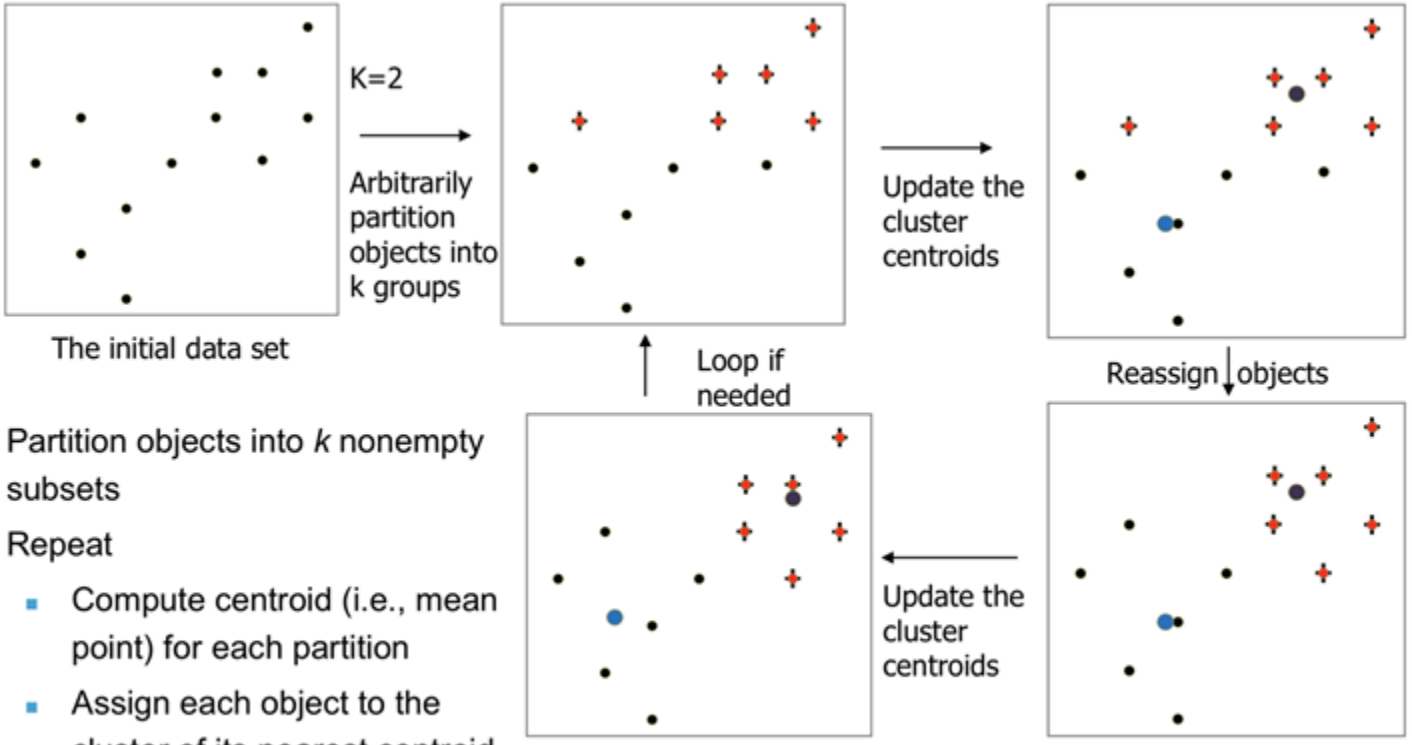
- Homework 3 due on 23:59pm, Nov 9, 2018
  - Submitted on CCLE and click “Submit for grading”
  - Must include: (1) Two python files: knn.py + neural_net.py (2) Two workbooks ipynb files with converted PDF files
- Midterm project report due on 23:59pm, Nov 12, 2018
  - Submit guidelines: https://docs.google.com/document/d/1xLeBU-n8nuMT6zhLyLL1NIGuu25SJU1OWl9eybdgjXg/edit?usp=sharing
- Midterm exam
  - Nov 14, In class (12:00-13:45)
  - Closed-book, one letter-size cheatsheet, one simple calculator
  - Keep every digital copy for convenience of grading → Clear and readable handwriting!
- No lecture on Nov 12 (Veteran Day)

Clustering

- Input / Output / Goal of clustering analysis
  - Large amount of unlabeled data in real life
- Difference between supervised learning and unsupervised learning
- Semi-supervised learning
- Typical clustering algorithm examples:
  - K-means
  - Hierarchical clustering
  - DB-SCAN
  - Mixture models



K-means



- Partition objects into *k* nonempty subsets
- Repeat
  - Compute centroid (i.e., mean point) for each partition
  - Assign each object to the cluster of its nearest centroid
- Until no change

UCLA

K-means

- Demo 1: http://stanford.edu/class/ee103/visualizations/kmeans/kmeans.html
- Demo 2: https://www.naftaliharris.com/blog/visualizing-k-means-clustering/
- Key idea of K-means algorithms:
  - Step 1: Partition into k non-empty subsets (select K points as initial centroids)
  - Step 2: Iteration: Update mean point and assign object to cluster again
  - Step 3: Stop when converge
- Partition-based clustering methods
- Can be considered as a special case of GMM

UCLA

K-means

- Q1: Will K-means converge?
- Q2: Will different initialization of K-means generate different clustering results?

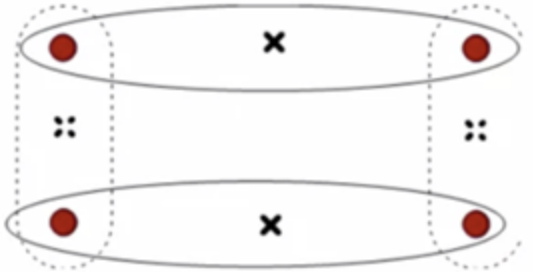
UCLA

K-means

- Q1: Will K-means converge?
- A1: Yes.

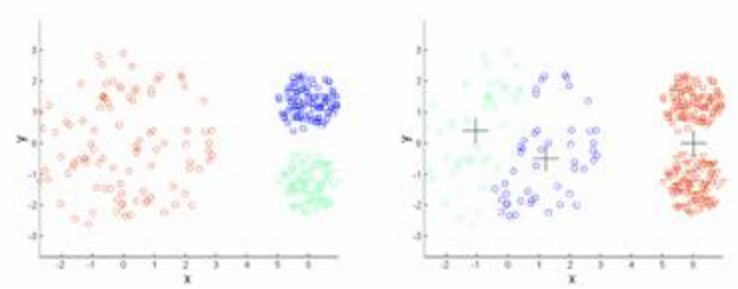
$$J = \sum_{j=1}^{k} \sum_{C(i)=j} d(x_i, c_j)^2$$

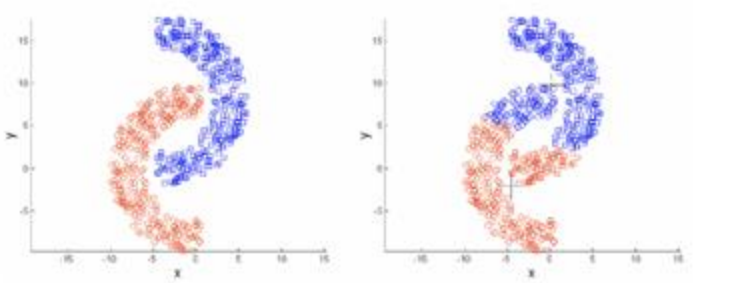
- Q2: Will different initialization of K-means generate different clustering results?
- A2: Yes. Initialization matters!



K-means: Discussion

- Efficiency: **O(tkn)** normally k,t are much smaller than n → efficient
- Can terminate at a local optima
- Need to specify k (or take time to find best k)
- Sensitive to noisy data and outliers
- Different sizes and variances
- Not suitable to discover clusters with non-convex shapes
  - Can K-medoids help?
- Many variants of K-means:
  - K-means++, Genetics K-means, etc.

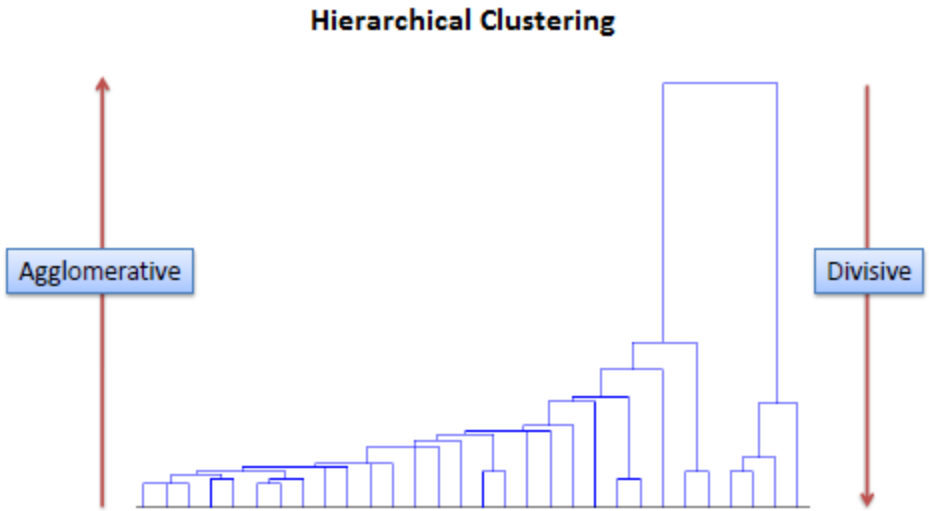




Hierarchical Clustering

- Method
  - Divisive (Top-down)
  - Agglomerative (Bottom-up)

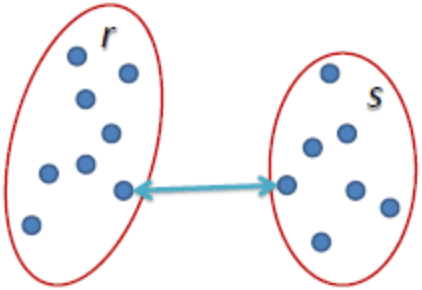
- Distance metrics
  - Single linkage
  - Complete linkage
  - Average linkage
  - Centroid
  - Medroid



UCLA

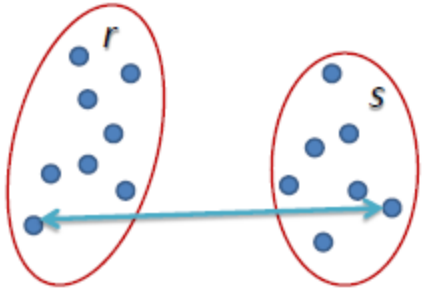
Hierarchical Clustering

- Single Linkage



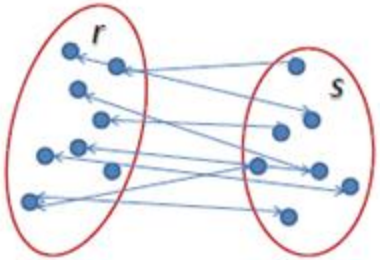
$$L(r,s) = \min(D(x_{ri}, x_{sj}))$$

- Complete Linkage



$$L(r,s) = \max(D(x_{ri}, x_{sj}))$$

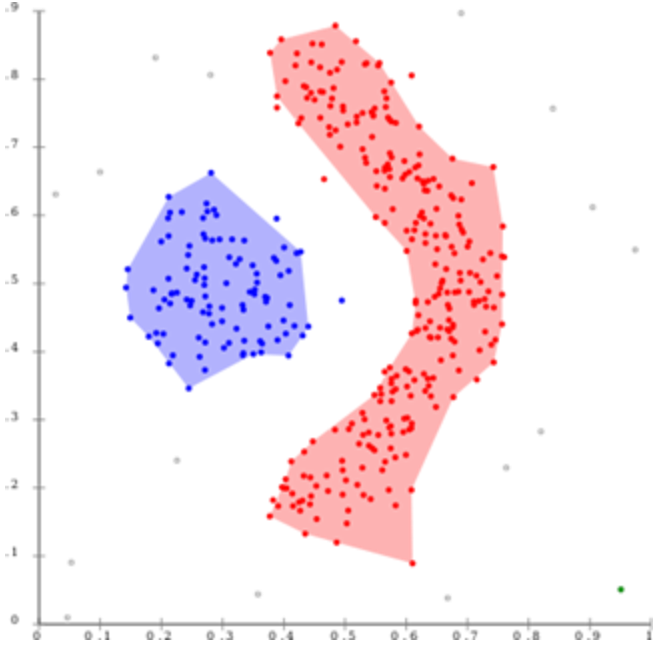
- Average Linkage

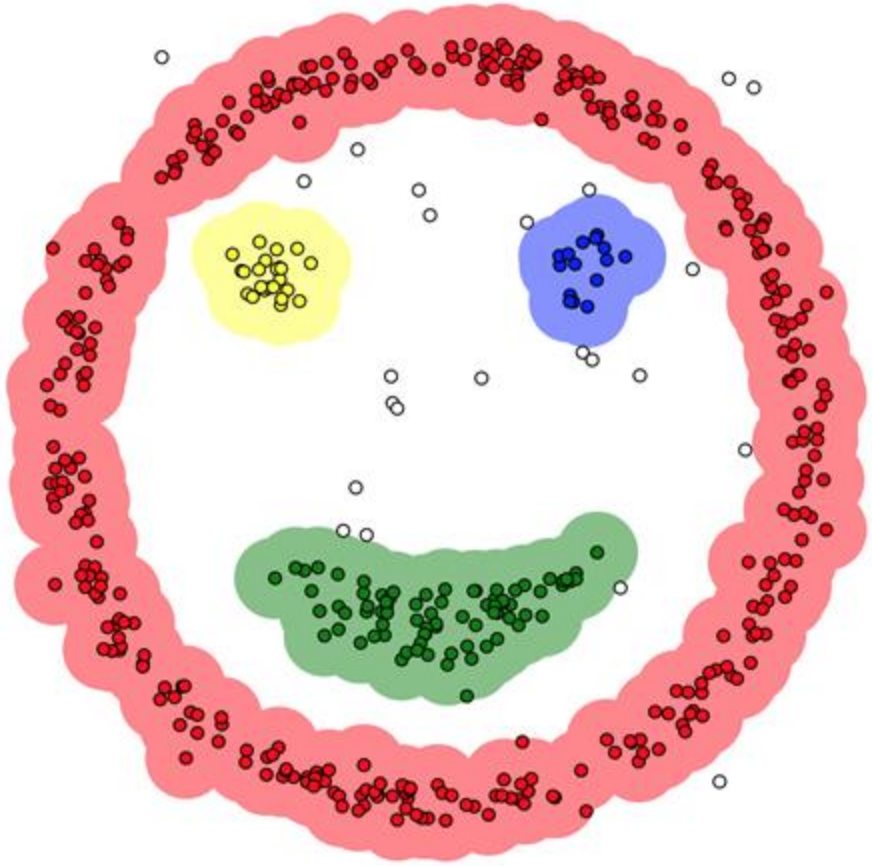


$$L(r,s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} D(x_{ri}, x_{sj})$$

DB-SCAN

- Density-based clustering method
- Demo
- Pros and Cons
  - has a notion of noise, and is robust to outliers
  - does not require one to specify the number of clusters in the data a priori, as opposed to k-means
  - can find arbitrarily shaped clusters
  - not entirely deterministic, depending on the order
  - cannot cluster data sets well with large differences in densities
- Complexity analysis



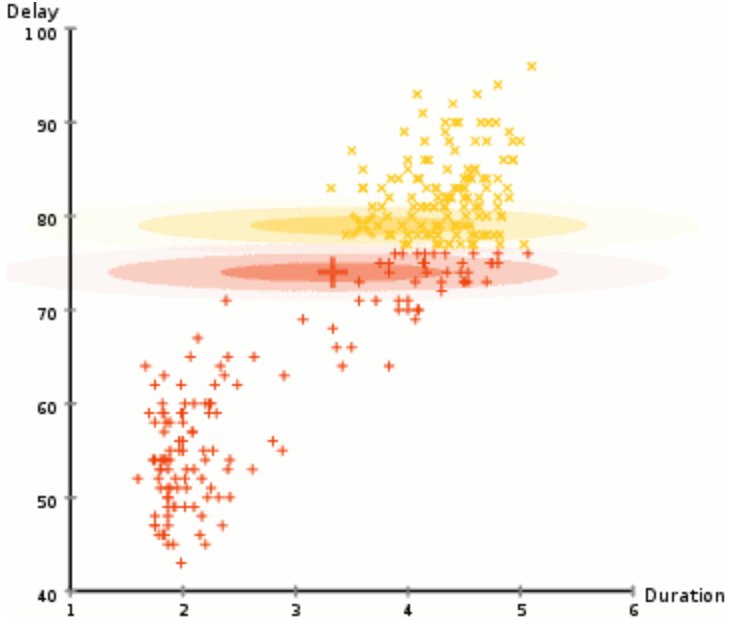


epsilon = 1.00
minPoints = 4

Restart

Gaussian Mixture Model

- Generative model vs Discriminative model
- EM algorithm
- More details on GMM



Example: Clustering Evaluation

Calculate:

Purity, precision, recall, F-measure, and normalized mutual information

ID	Conference Name	Ground Truth Label	Algorithm output Label
1	IJCAI	3	2
2	AAAI	3	2
3	ICDE	1	3
4	VLDB	1	3
5	SIGMOD	1	3
6	SIGIR	4	4
7	ICML	3	2
8	NIPS	3	2
9	CIKM	4	3
10	KDD	2	1
11	WWW	4	4
12	PAKDD	2	1
13	PODS	1	3
14	ICDM	2	1
15	ECML	3	2
16	PKDD	2	1
17	EDBT	1	2
18	SDM	2	1
19	ECIR	4	4
20	WSDM	4	4

UCLA

Midterm Reminder

- **Time:** 12:00-13:45, Nov 14
- **Place:** In Class
- **Questions:** True/False, Conceptual questions, Computation questions
- **Policy:**
  - Honor code (No discussion during exam).
  - Closed book exam, but you can take a “reference sheet” of letter size.
  - You can bring a simple calculator.
  - Make sure your handwriting readable and clear enough and write down your answer in the “correct” area. → We will use the digital copy for grading.
  - Keep your answer concise.

UCLA

Midterm Topics

Task	Vector data
Classification	Logistic Regression; Decision Tree; KNN; SVM; Neural Networks
Clustering	K-means; Hierarchical clustering; DBSCAN; Mixture Models
Prediction	Linear Regression

UCLA

True or False

Sample Questions

- You are only required to write down “T” or “F” before each statement. No need to justify (subject to change).
- Insights and details about learned algorithms and techniques.

UCLA

Examples: True or False

Sample Questions

[?] Logistic regression is a linear classifier.

[?] Normalization of the features (e.g., z-score) in linear regression will affect the weights learned from closed-form solution.

[?] Decision tree can handle both numerical and categorical features.

[?] In K Nearest Neighbor classifier, a bigger K will more likely cause overfitting in the training dataset.

[?] Linear regression is a supervised learning model while hierarchical clustering is unsupervised learning model.

UCLA

Conceptual Questions

Sample Questions

Give your conclusion or results and also need to justify (in most cases).

- **Example 1:** What is the time complexity of DBSCAN? Why?
- **Example 2:** What are the conditions of stopping partitioning the training examples when constructing a decision tree?

UCLA

Computation Questions

Sample Questions

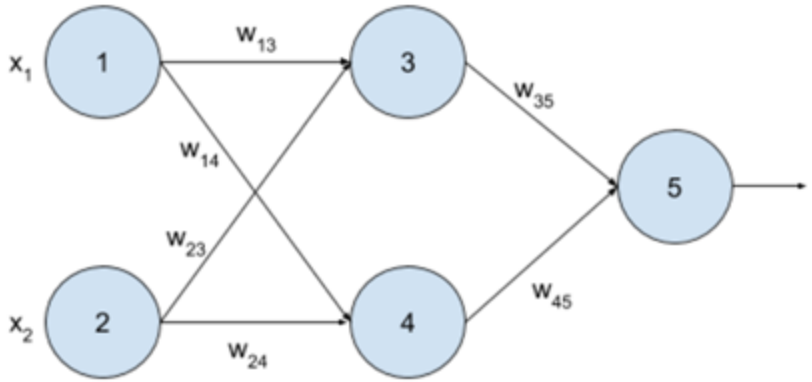
For example,

- Information gain computation in decision tree
- SVM problem in HW2
- Neural network forward computation
- Classification evaluation

Example: Computation Questions

Sample Questions

- Given the following multilayer neural network, a training data point x = (x_1=0, x_2=1), and the target value T=1, please calculate weights and bias after 1 iteration of backpropagation algorithm (show your calculations and fill out the empty tables given below). The learning rate =0.8. The initial weights and bias are in the following table.



w_{13}	w_{14}	w_{23}	w_{24}	w_{35}	w_{45}	3	4	5
-0.3	0.2	0.4	-0.1	-0.2	-0.3	0.2	-0.4	0.1

UCLA

Examples

- Given the following multilayer neural network, a training data point $x=(x_1=0,x_2=1)$, and the target value T=1, please calculate weights and bias after 1 iteration of backpropagation algorithm (show your calculations and fill out the empty tables given below). The learning rate =0.8. The initial weights and bias are in the following table.

w_{13}	w_{14}	w_{23}	w_{24}	w_{35}	w_{45}	3	4	5
-0.3	0.2	0.4	-0.1	-0.2	-0.3	0.2	-0.4	0.1

Net Input and Output Calculations

Unit, j	Net Input, I_j	Output, O_j
3	- 0.3 (0) + 0.4 (1) + 0.2 = 0.6	0.6457
4	0.2 (0) - 0.1 (1) – 0.4 = - 0.5	0.3775
5	- 0.2 (0.6457) – 0.3 (0.3775) + 0.1 = - 0.14239	0.4645

Calculation of the error at each node

Pay attention to whether it is **err** or **derivative**.

Unit, j	Err_j
5	0.4645 (1 - 0.4645) (1 – 0.4645) = 0.1332
4	0.3775 (1 – 0.3775) (0.1332) (- 0.3) = - 0.0094
3	0.6457 (1 – 0.6457) (0.1332) (- 0.2) = - 0.0061

Calculations for weight and bias updating

Pay attention to the sign here! If **err**, +; If **derivative**, - (~SGD).

Weight or Bias	New Value
w_{35}	- 0.2 + 0.8 (0.1332) (0.6457) = - 0.1312
w_{45}	- 0.3 + 0.8 (0.1332) (0.3775) = - 0.2598
w_{13}	- 0.3 + 0.8 (- 0.0061) (0) = - 0.3
w_{14}	0.2 + 0.8 (- 0.0094) (0) = 0.2
w_{23}	0.4 + 0.8 (- 0.0061) (1) = 0.3951
w_{24}	- 0.1 + 0.8 (- 0.0094) (1) = - 0.1075
θ_5	0.1 + 0.8 (0.1332) = 0.2066
θ_4	- 0.4 + 0.8 (- 0.0094) = - 0.4075
θ_3	0.2 + 0.8 (- 0.0061) = 0.1951

UCLA

Thank you!

Q & A