Final Review

Lecture Overview

1. Final Structure
2. Naïve Bayes Classification
3. SVM Classifier
4. Hierarchical Clustering
5. K-means
6. DBSCAN
7. Cophenetic Correlation

Final

- Total comes to 120 pts. However, max you can earn is capped at 100.
- Same set of answer sheets for graduate and for undergraduate – however questions are marked as follows:
  - A (e.g., ‘problem 1A’) – for both graduate and undergraduate
  - G – (e.g. ‘problem 2G’) for graduate

  Please pay attention and avoid confusion.
- 5 questions UG + 1 question G
- Computationally intensive - you should use Wolfram alpha or calculator.
- You MUST give ALL intermediate results whenever asked for. No intermediate results no credit

Final - continued

- Credit for each question clearly marked. Partial credit possible, also marked.
  - Credit is based on undergraduate credit
  - Undergrad credit for a UG question is given in brackets e.g. [30] means that a question gives undergraduate student 30 pts.
  - Graduate credit for UG is based on undergraduate credit multiplied by a common multiplier 5/6, e.g. [30] means that grad students get for this question 30*5/6=25 points. Same multiplier applies to all partial points
  - Grad credit for grad only questions (G) credit is given in brackets so [25] on a question marked G means grad students get 25 pts for this question

Final - continued

Each question has a close sibling in the HW and/or earlier exams – we'll see later today.

Points for UG problems are given for undergrads. For grad students multiplier for these problems is 4/5

1. Naïve Bayes Classification
2. SVM classification
3. Hierarchical clustering

Twists:

- the data points are binary 7-dimensional vectors (i.e. each entry is 0 or 1);
- the proximity between data points is not given in the problem – you need to compute similarity/distance matrix first.
- Metric in $\mathbb{B}^7$ is Hamming distance

Final - continued

4. Clustering K-means. Given a set of data points. Explain how k-means with a given k will cluster it.
5. A picture shown in the problem, is given to the DBSCAN algorithm along with radius eps and core threshold minpts. Show what would be DBSCAN's clusterization result. How does it change when we change the radius?
6. G only; Evaluation of hierarchical clustering. Compute cophenetic distance table and cophenetic correlation coefficient.

Lecture Overview

1. Final Structure
2. Naïve Bayes Classification
3. SVM Classifier
4. Hierarchical Clustering
5. K-means
6. DBSCAN
7. Cophenetic Correlation

Q1: 'Buy Computer' Classifier

Given the training dataset in table 1 below (Buy Computer data), predict the class of the following new record using Nave Bayes classication:

<age 30; income=medium; student=yes; credit-rating=fair>

RID	age	income	student	credit_rating	Class: buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31 . . . 40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31 . . . 40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31 . . . 40	medium	no	excellent	yes
13	31 . . . 40	high	yes	fair	yes
14	>40	medium	no	excellent	no

Q1: Record Classification by Naïve Bayes

Priors (frequencies) of classes are:

$$P(yes) = \frac{9}{14} = 0{,}643 \qquad P(no) = \frac{5}{14} = 0.357$$

For the record R=<age 30; income = medium; student = yes; credit rating = fair> we also need $P(age < 30|C)$, $P(inc = m|C)$, $P(st = Y|C)$, $P(CR = Fair|C)$ for each class C.

$$P(age < 30|Yes) = \frac{2}{9} = 0.222 \qquad P(age < 30|No) = \frac{3}{5} = 0.6$$

$$P(inc = m|Yes) = \frac{4}{9} = 0.444 \qquad P(inc = m|No) = \frac{2}{5} = 0.4$$

$$P(st = Y|C) = \frac{6}{9} = 0.667 \qquad P(st = Y|No) = \frac{1}{5} = 0.2$$

$$P(CR = Fair|Yes) = \frac{6}{9} = 0.667 \qquad P(CR = Fair|NO) = \frac{2}{5} = 0.4$$

$$\frac{P(Yes|R)}{P(No|R)} = \frac{0.222 \times 0.444 \times 0.667 \times 0.667 \times 0.643}{0.6 \times 0.4 \times 0.2 \times 0.4 \times 0.357} > 1,$$

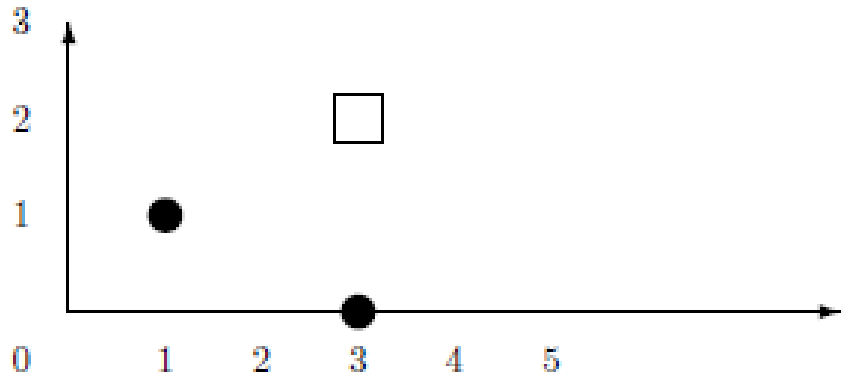
So the record is classified as 'yes'

Lecture Overview

1. Final Structure
2. Naïve Bayes Classification
3. SVM Classifier
4. Hierarchical Clustering
5. K-means
6. DBSCAN
7. Cophenetic Correlation

Q2: 2-dimensional SVM

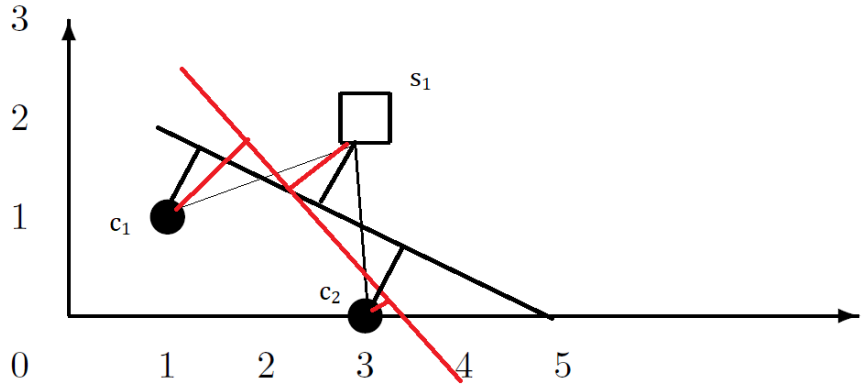
- **Suppose we are given tiny two class data set (dots and squares) shown on fig below Build SVM for this set.**



- Determine supporting vectors and prove that they are indeed supporting vectors
- Give the construction of your hyperplane based on your supporting vectors

Q2: SVM Support Vectors

- Claim. G 3 points in $\mathbb{R}^2$ that do not lie on a line. Then the equidistant separation plane maximizes the margin between positive and negative examples.
- Proof. Let $\vec{p}_1, \vec{p}_2;\ \vec{p}_3$ be these orthogonal margin vectors for $\vec{c}_1, \vec{c}_2$ and $\vec{s}_3$ resp. The equidistant plane goes though the midpoints of vectors $\vec{c}_1 - \vec{s}_3$, and $\vec{c}_2 - \vec{s}_3$. Alternative separating plane $\vec{c}_1 - \vec{s}_3$, and $\vec{c}_2 - \vec{s}_3$ at intersection points at least one of which will be closer one of the $\vec{c}_1, \vec{c}_2$ and $\vec{s}_3$ than midpoints of vectors $\vec{c}_1 - \vec{s}_3$, and $\vec{c}_2 - \vec{s}_3$. Say it is closer to c_2, but then it should also intersect $\vec{p}_2$ (orthogonal to equidistant plane). Thus the orthogonal to this alternative plane is going to be shorter than equidistant margin because orthogonal has shortest distance to plane.



Q2: SVM Support Vectors

- So we have two data points of class +='circle': $\binom{3}{0}$ and $\binom{1}{1}$ and one data point of class -='square': $\binom{3}{2}$ that are not on the line, so they all are the support vectors.
- So if they are support vectors then they should satisfy the following system of equations:

$$\begin{cases} \binom{w_1}{w_2} \cdot \binom{3}{0} + b = 1 \\ \binom{w_1}{w_2} \cdot \binom{1}{1} + b = 1 \\ \binom{w_1}{w_2} \cdot \binom{3}{2} + b = -1 \end{cases}$$

Q2: Classifying Hyperplane

- If these vectors are support vectors then they should satisfy the following system of equations:

$$\begin{cases} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 0 \end{pmatrix} + b = 1 \\ \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} + b = 1 \\ \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 2 \end{pmatrix} + b = -1 \end{cases}$$

The corresponding augmented matrix is

$$\begin{pmatrix} 3 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 3 & 2 & 1 & -1 \end{pmatrix}$$

Which gives the solution $\overline{w} = \begin{pmatrix} -\frac{1}{2} \\ -1 \end{pmatrix}$ and $b = \frac{5}{2}$ therefore the hyperplane (line) $[\overline{w}: b] = -\frac{1}{2}x_1 - 1x_2 + \frac{5}{2} = 0$ or equivalently

$$x_2 = -0.5x_1 + 2.5$$

Lecture Overview

1. Final Structure
2. Naïve Bayes Classification
3. SVM Classifier
4. Hierarchical Clustering
5. K-means
6. DBSCAN
7. Cophenetic Correlation

Q3: Similarities and Distances

- For hierarchical clustering you could be asked to use either distances or similarities
- If distances then closest two points are at distance that is closest to 0; If similarities then closest are the points similarity between which is closest to 1
- Problem will start with points and you could be asked to use unorthodox distances or similarities
- It could be unusual measures that you would need to compute before doing hierarchical clustering, for example
  - Given points in $\binom{2}{0}, \binom{1}{1}, \binom{0}{1} \in \mathbb{R}^2$ use cosine similarity $\cos(a,b) = \frac{a \cdot b}{\|a\| \times \|b\|}$, so $\cos\left(\binom{2}{0}, \binom{1}{1}\right) = \frac{\sqrt{2}}{2}$; $\cos\left(\binom{2}{0}, \binom{0}{1}\right) = 0$; $\cos\left(\binom{1}{1}, \binom{0}{1}\right) = \frac{\sqrt{2}}{2}$
  - Given the following data points that are 5-dimensional Boolean data, use Hamming distance (i.e. the number of mismatches in respective positions): $A = (1,0,1,1,0), B = (1,1,0,1,1), C = (0,0,1,1,0)$.

    Hamming distances are $h(A,B) = 3, h(A,C) = 1, h(B,C) = 4$

Q3:Aggloomerative Clustering

Use the similarity matrix in Table 1 to perform single and complete link hierarchical clustering. Show your results by drawing dendrogram. The dendrogram should clearly show the order in which the points are merged

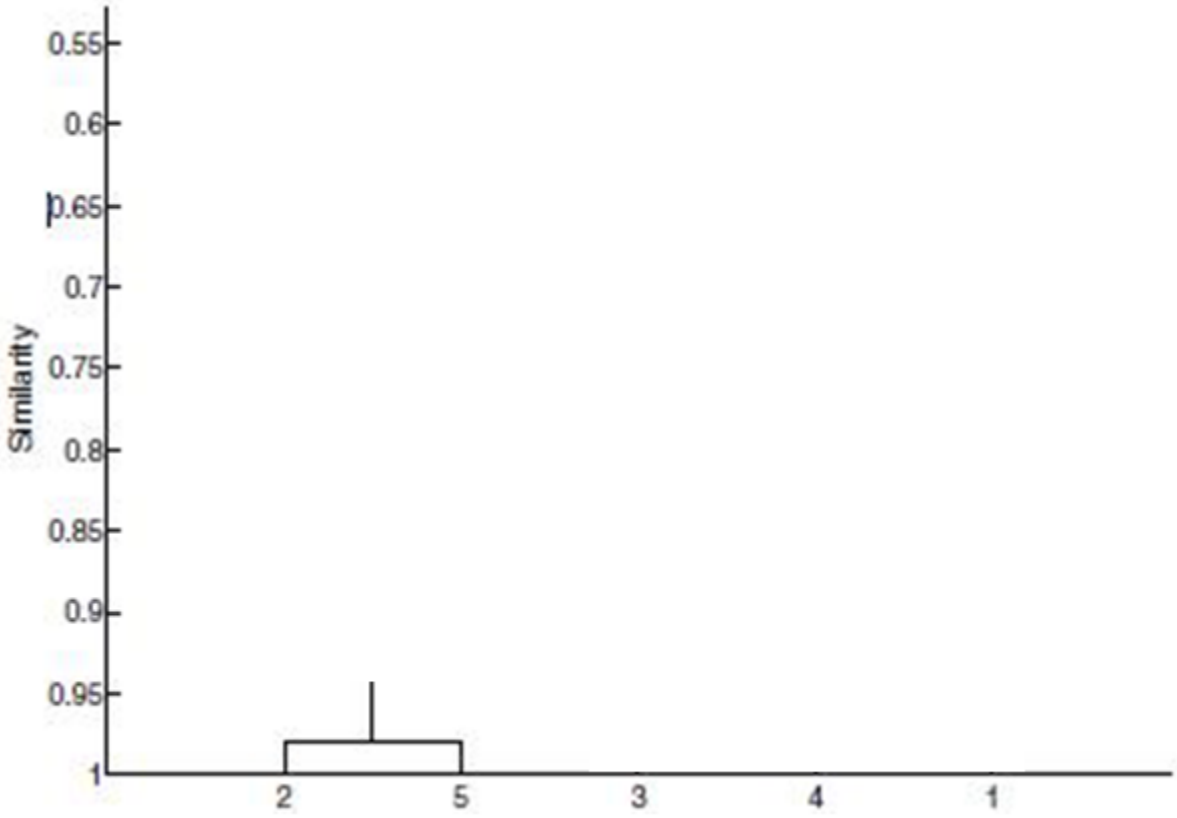
	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

Table 1

Q3: Single Link

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

	p1	p2	p3	p4	p5
p1	1.00	0.35	0.41	0.55	0.35
p2	0.35	1.00	0.85	0.76	1.00
p3	0.41	0.85	1.00	0.44	0.85
p4	0.55	0.76	0.44	1.00	0.76
p5	0.35	1.00	0.85	0.76	1.00

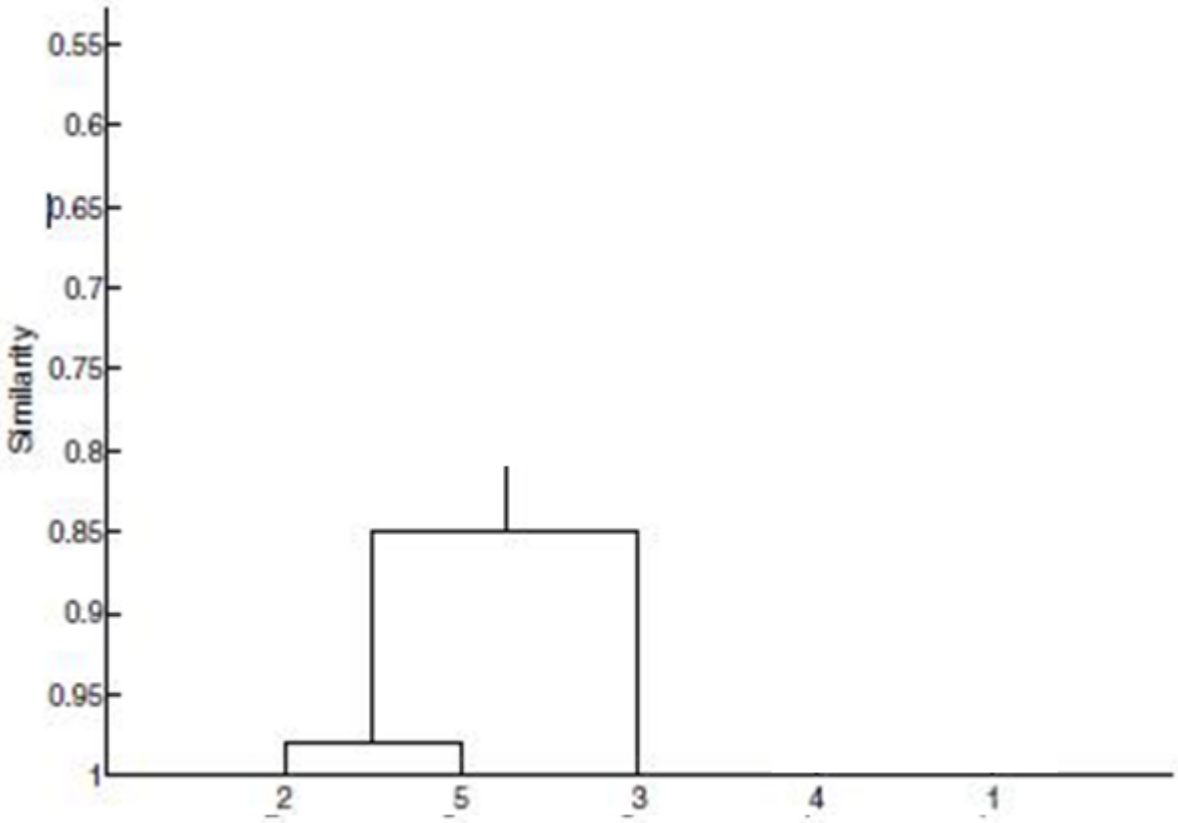


Q3: Single Link

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

	p1	p2	p3	p4	p5
p1	1.00	0.35	0.41	0.55	0.35
p2	0.35	1.00	0.85	0.76	1.00
p3	0.41	0.85	1.00	0.44	0.85
p4	0.55	0.76	0.44	1.00	0.76
p5	0.35	1.00	0.85	0.76	1.00

	p1	p2	p3	p4	p5
p1	1.00	0.41	0.41	0.55	0.41
p2	0.41	1.00	1.00	0.76	1.00
p3	0.41	1.00	1.00	0.76	1.00
p4	0.55	0.76	0.76	1.00	0.76
p5	0.41	1.00	1.00	0.76	1.00

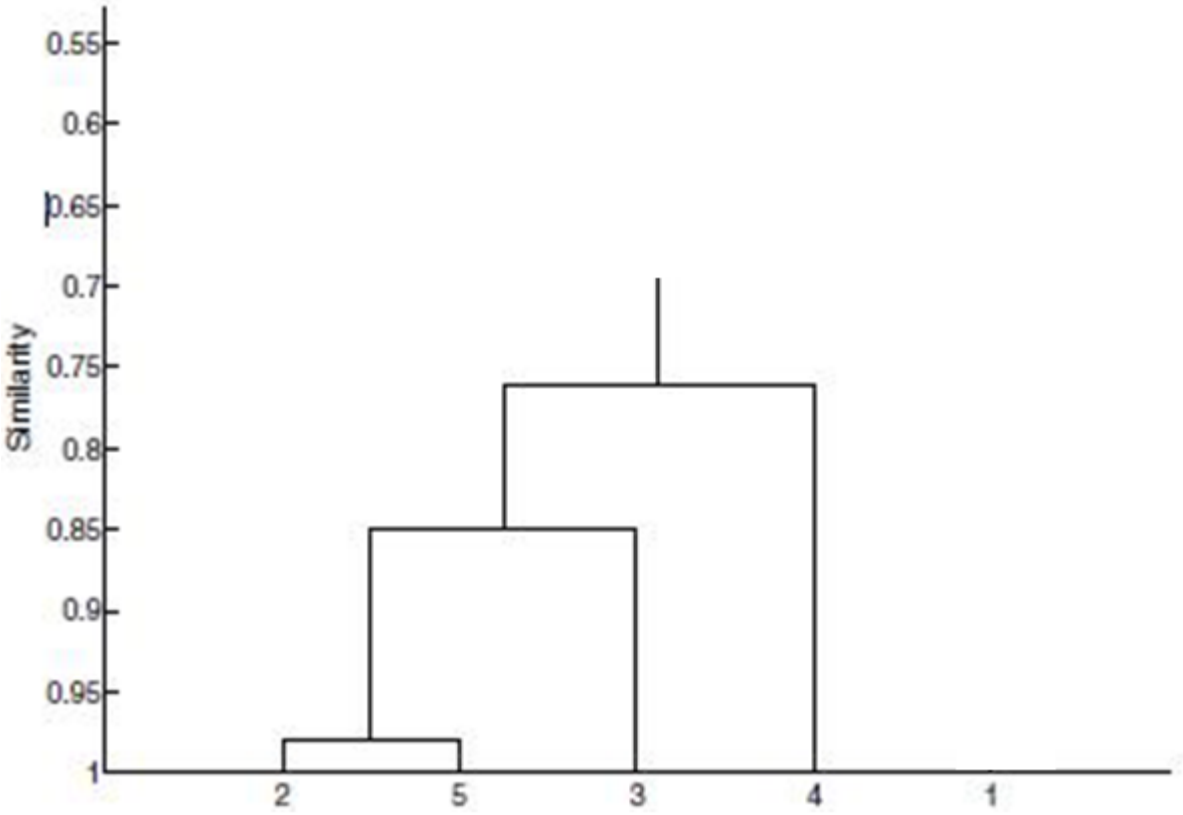


Q3: Single Link

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

	p1	p2	p3	p4	p5
p1	1.00	0.35	0.41	0.55	0.35
p2	0.35	1.00	0.85	0.76	1.00
p3	0.41	0.85	1.00	0.44	0.85
p4	0.55	0.76	0.44	1.00	0.76
p5	0.35	1.00	0.85	0.76	1.00

	p1	p2	p3	p4	p5
p1	1.00	0.41	0.41	0.55	0.41
p2	0.41	1.00	1.00	0.76	1.00
p3	0.41	1.00	1.00	0.76	1.00
p4	0.55	0.76	0.76	1.00	0.76
p5	0.41	1.00	1.00	0.76	1.00

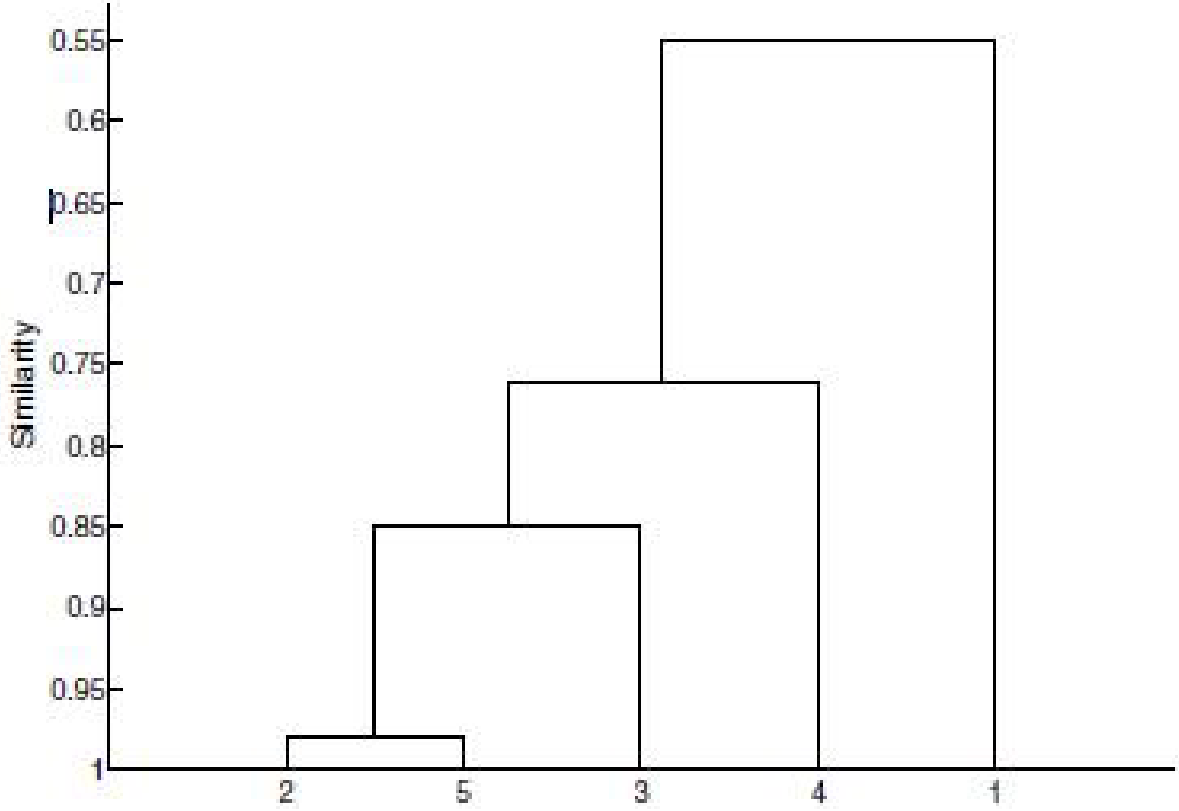


	p1	p2	p3	p4	p5
p1	1.00	0.55	0.55	0.55	0.55
p2	0.55	1.00	1.00	1.00	1.00
p3	0.55	1.00	1.00	1.00	1.00
p4	0.55	1.00	1.00	1.00	1.00
p5	0.55	1.00	1.00	1.00	1.00

Q3: Single Link

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

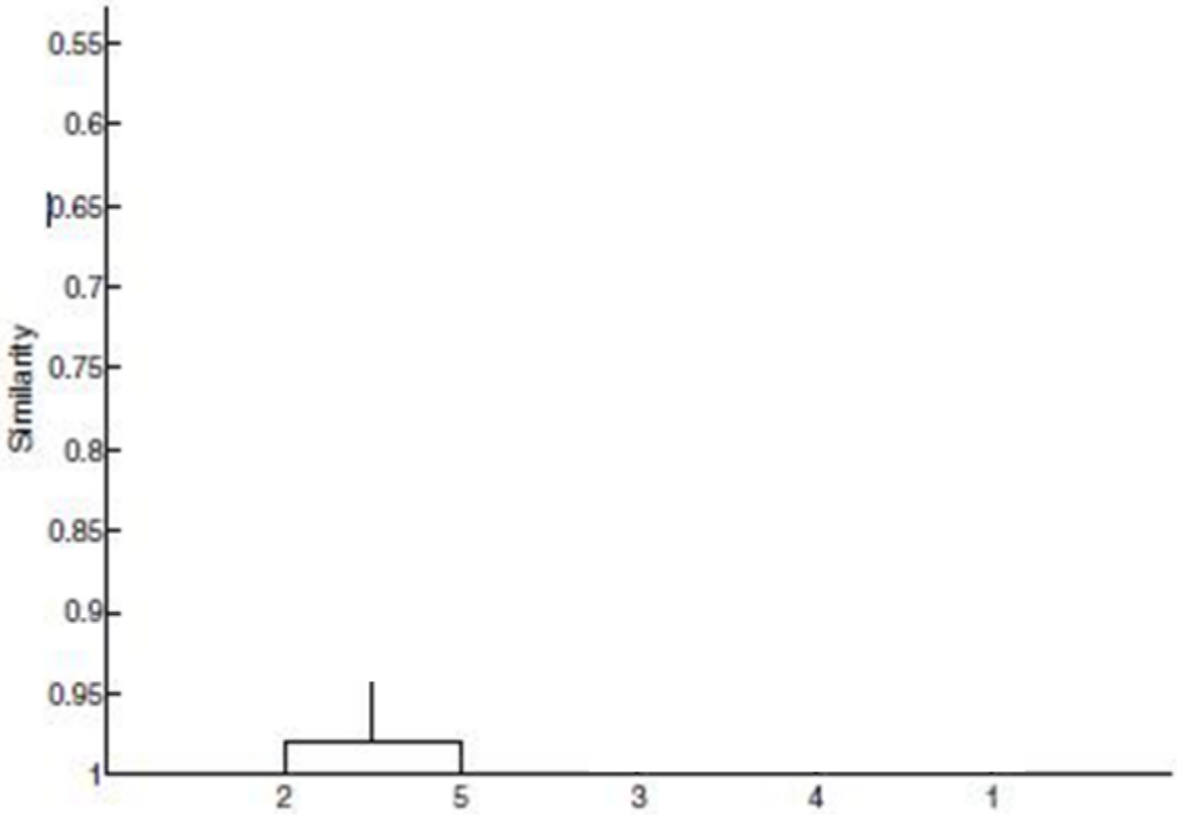
Table 1



Q3: Complete Link

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.10
p2	0.10	1.00	0.64	0.47	1.00
p3	0.41	0.64	1.00	0.44	0.64
p4	0.55	0.47	0.44	1.00	0.47
p5	0.10	1.00	0.64	0.47	1.00

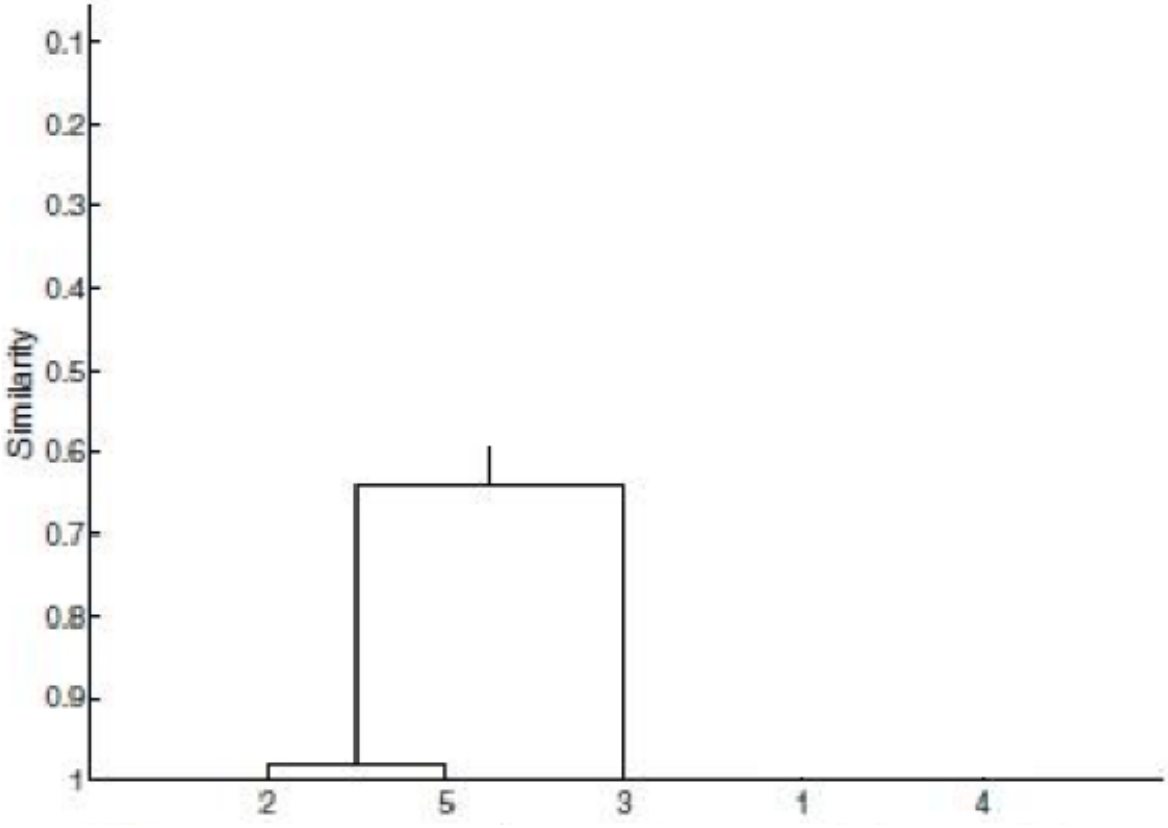


Q3: Complete Link

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.10
p2	0.10	1.00	0.64	0.47	1.00
p3	0.41	0.64	1.00	0.44	0.64
p4	0.55	0.47	0.44	1.00	0.47
p5	0.10	1.00	0.64	0.47	1.00

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.10	0.55	0.10
p2	0.10	1.00	1.00	0.44	1.00
p3	0.10	1.00	1.00	0.44	1.00
p4	0.55	0.44	0.44	1.00	0.44
p5	0.10	1.00	1.00	0.44	1.00

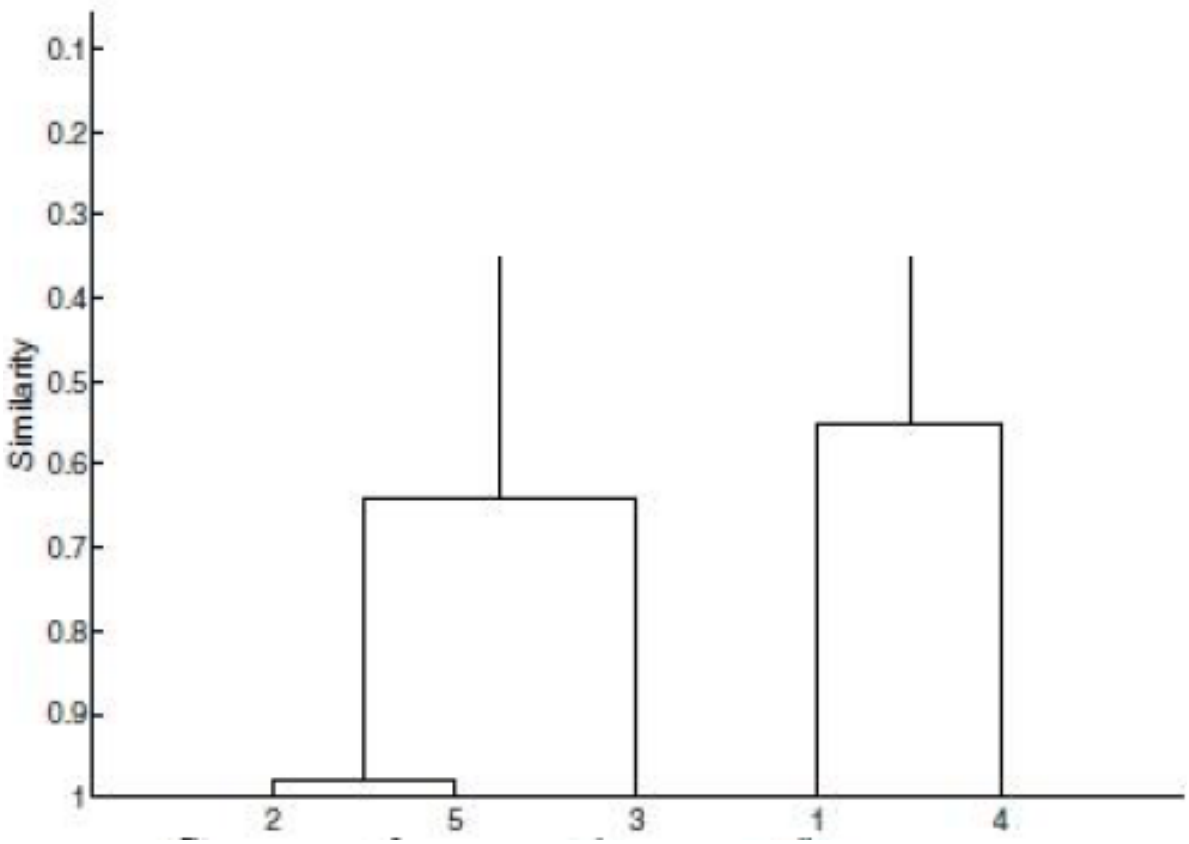


Q3: Complete Link

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.10
p2	0.10	1.00	0.64	0.47	1.00
p3	0.41	0.64	1.00	0.44	0.64
p4	0.55	0.47	0.44	1.00	0.47
p5	0.10	1.00	0.64	0.47	1.00

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.10	0.55	0.10
p2	0.10	1.00	1.00	0.44	1.00
p3	0.10	1.00	1.00	0.44	1.00
p4	0.55	0.44	0.44	1.00	0.44
p5	0.10	1.00	1.00	0.44	1.00

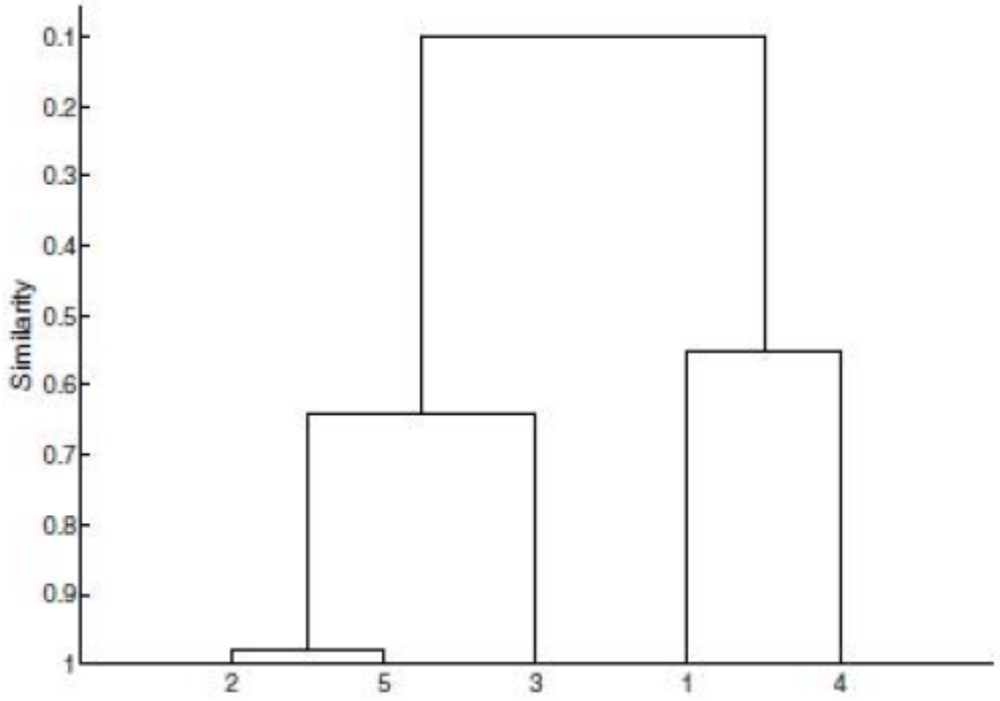


	p1	p2	p3	p4	p5
p1	1.00	0.10	0.10	1.00	0.10
p2	0.10	1.00	1.00	0.10	1.00
p3	0.10	1.00	1.00	0.10	1.00
p4	1.00	0.10	0.10	1.00	0.10
p5	0.10	1.00	1.00	0.10	1.00

Q3: Complete Link

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

Table 1

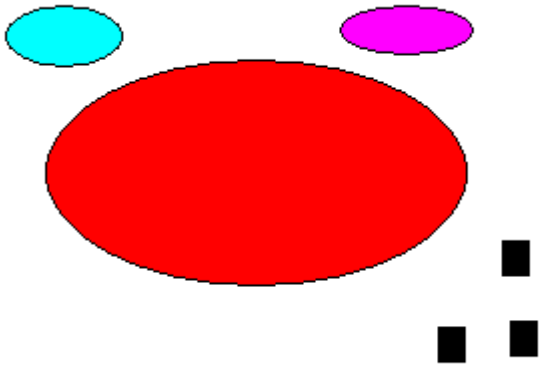


Lecture Overview

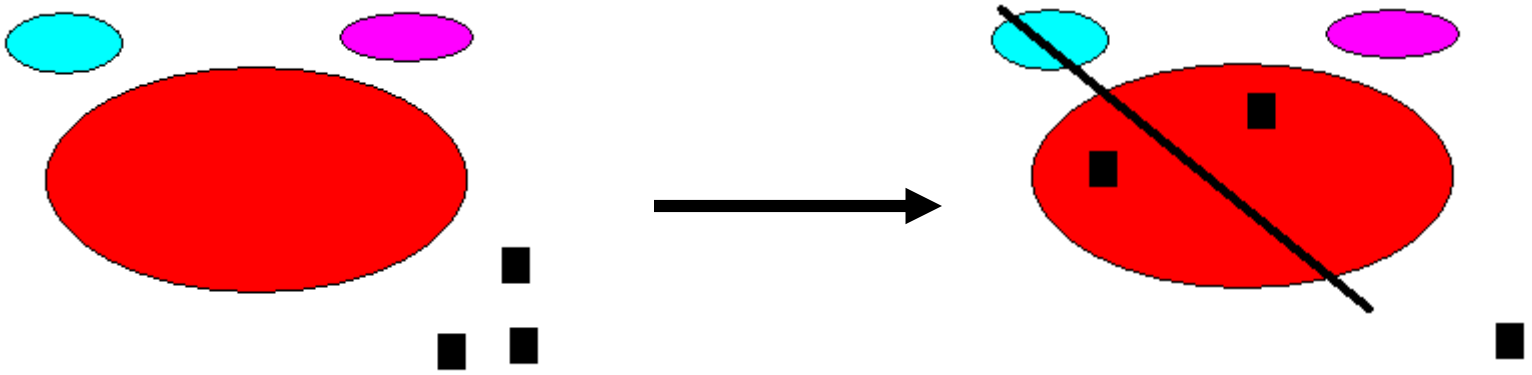
1. Final Structure
2. Naïve Bayes Classification
3. SVM Classifier
4. Hierarchical Clustering
5. K-means
6. DBSCAN
7. Cophenetic Correlation

Q4: Graphic Explanation of K-means

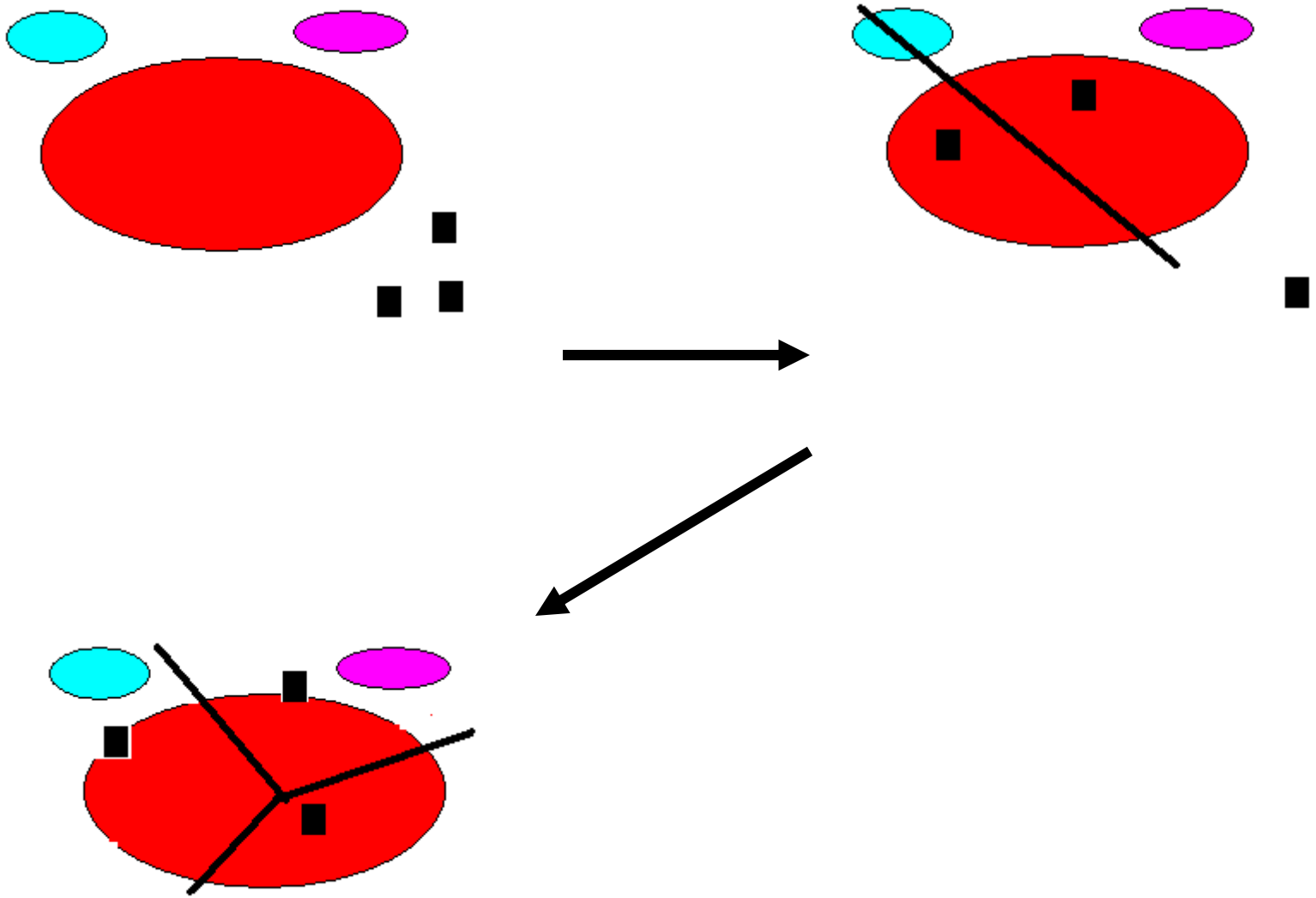
Given a graphic dataset. Explain how k-means with a given k cluster it.



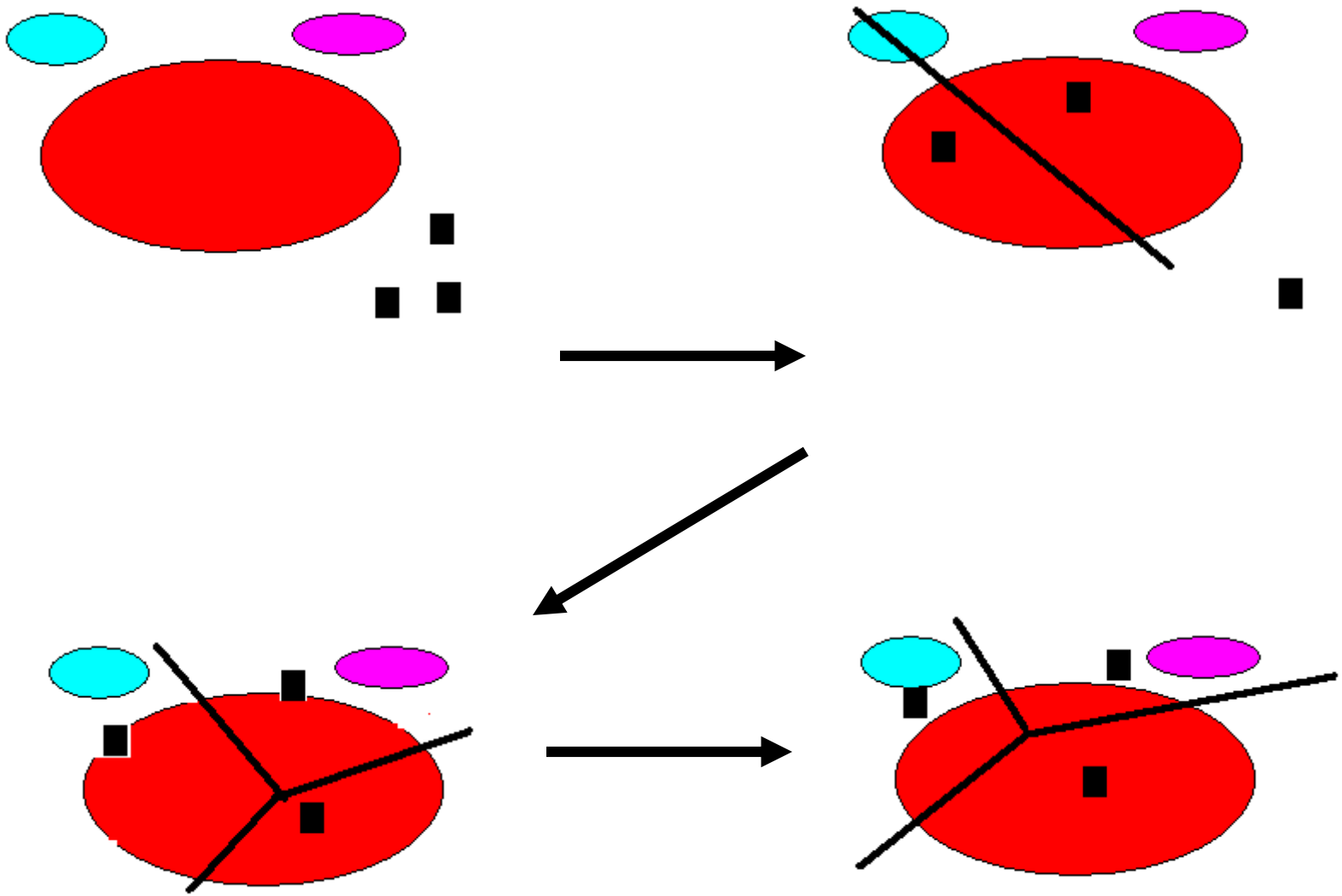
Q4: Graphic Explanation of K-means cont.



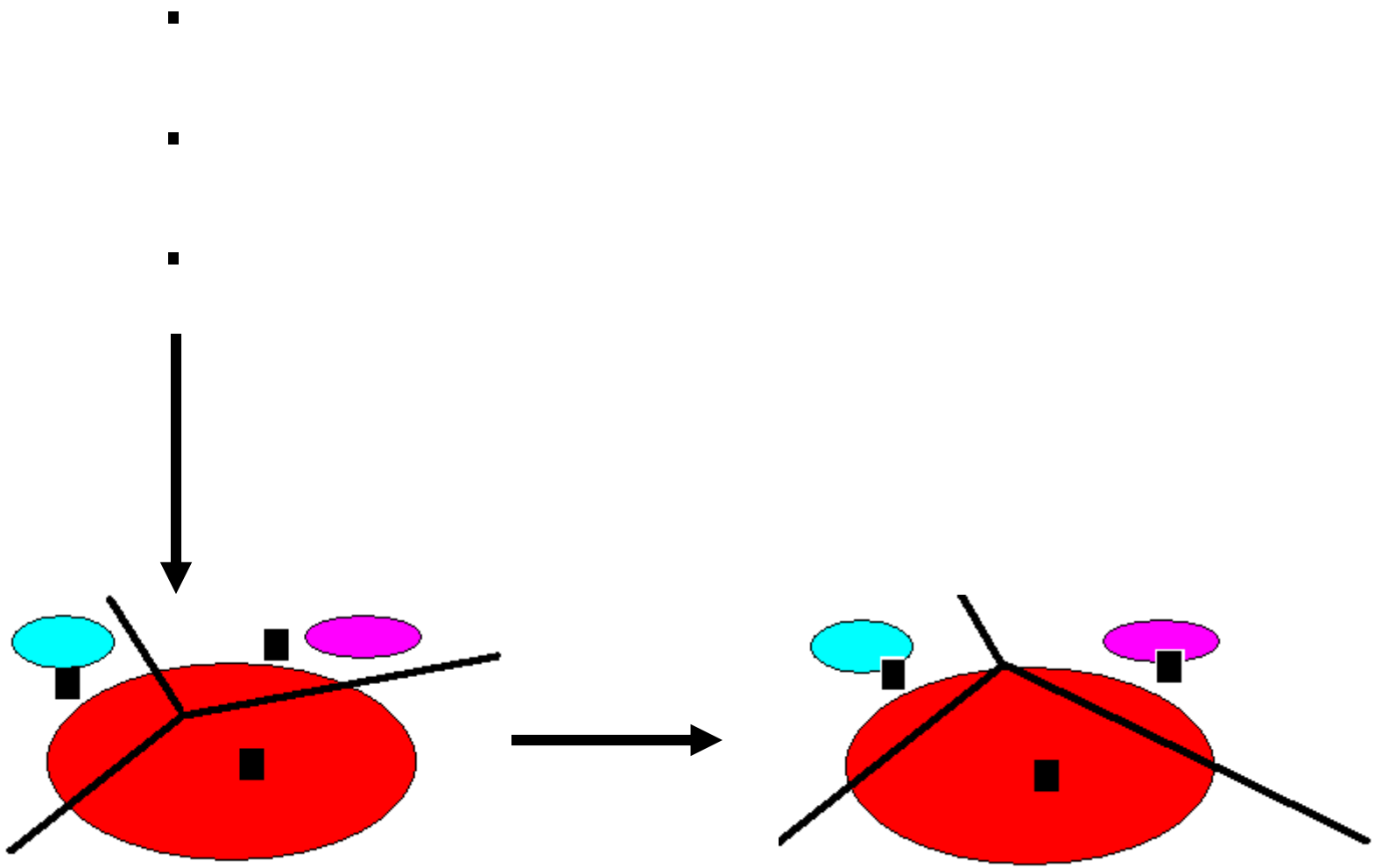
Q4: Graphic Explanation of K-means cont..



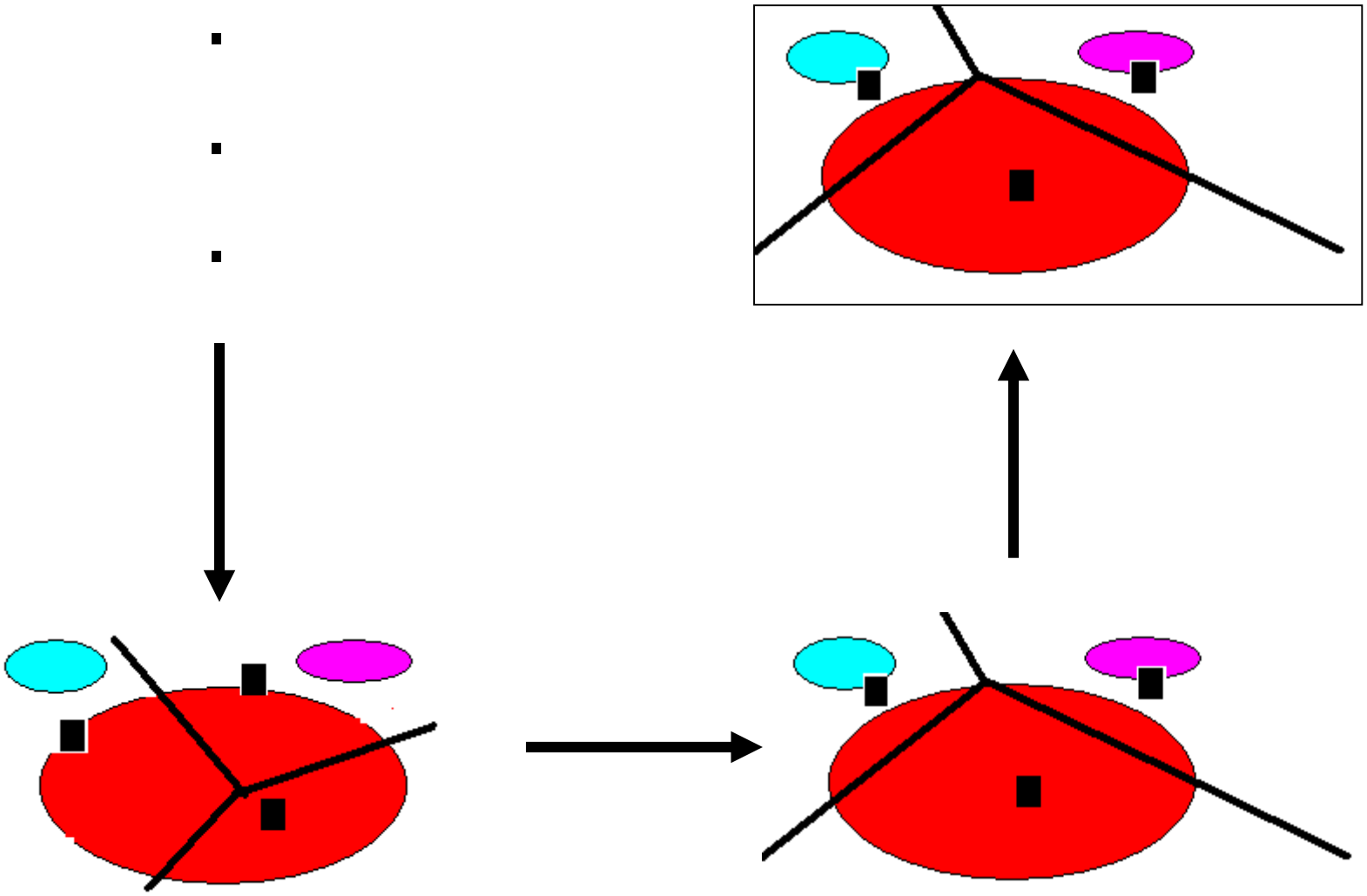
Q4: Graphic Explanation of K-means cont.



Q4: Graphic Explanation of K-means cont.



Q4: Graphic Explanation of K-means cont.

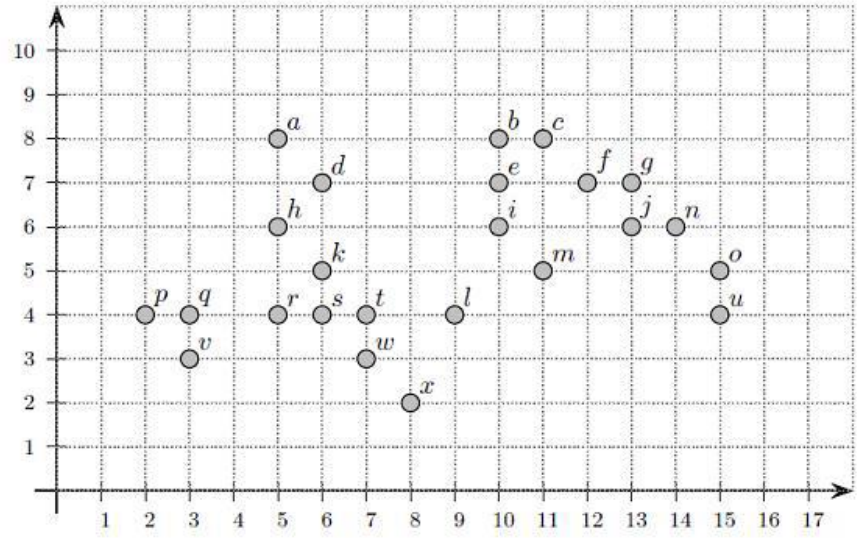


Lecture Overview

1. Final Structure
2. Naïve Bayes Classification
3. SVM Classifier
4. Hierarchical Clustering
5. K-means
6. DBSCAN
7. Cophenetic Correlation

Q5: Density-based Clusterization

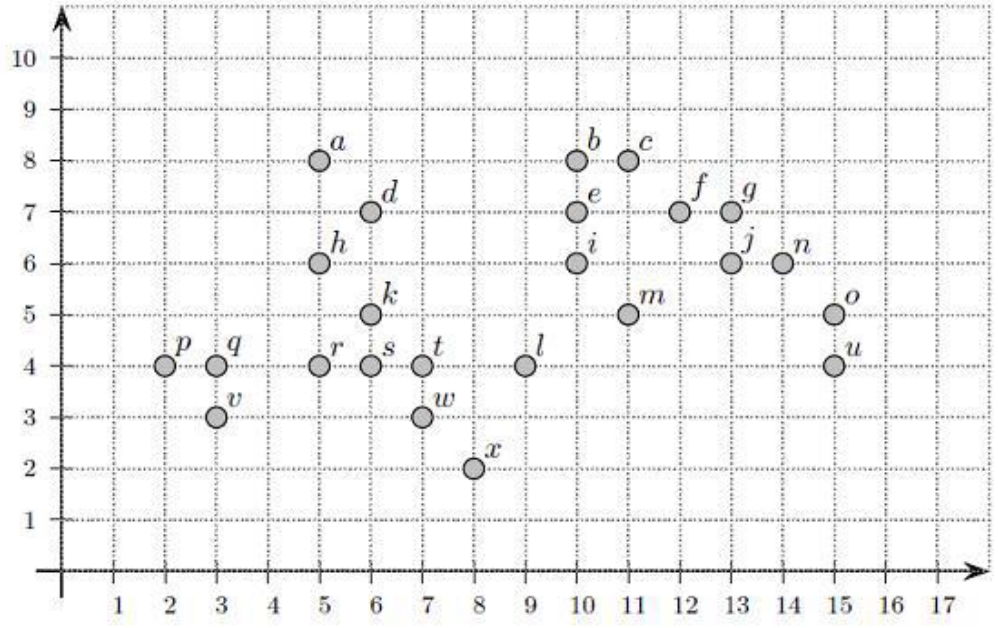
- Consider figure below. Answer the following questions related to DBSCAN. Assume that we use the Euclidean distance between points in DBSCAN, and that radius is eps = 1.5 and threshold for a point to be in the core is to have at least minpts = 3 neighbors within the radius



Q5: DBSCAN Core Points

List all core points.

- Solution:
- All points except $x; l; a; m; u$ are core points.

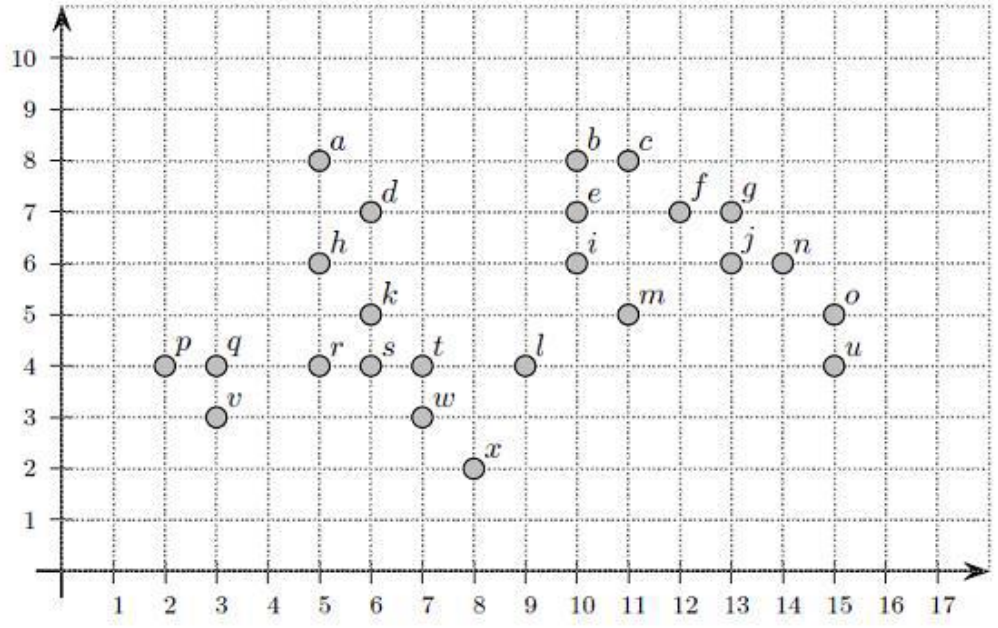


We say that a point x is directly density reachable from another point y, if x belongs to the neighborhood $N_{eps}(y)$ (i.e. x is in the ball of eps-radius around y) and y is a core point. Is a directly density-reachable from d?

- Yes it is since d is a core point and a is in its radius 1.5 ball

Q5: DBSCAN Density Reachability

We say that x is density reachable from y if there is a chain of points $x = c_1, \ldots, c_k = y$ such that c_i is directly reachable from c_{i-1} for all i. Is density-reachable a symmetric relationship, i.e., if x is density-reachable from y,

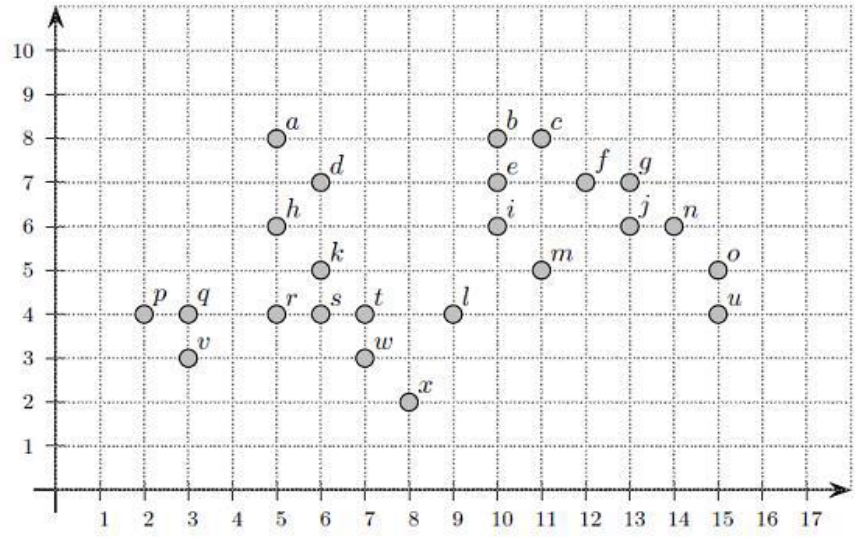


does it imply that y is density-reachable from x? Why or why not?

No, it is not. Even-'directly density-reachable' is not a symmetric relation. For example u is directly density reachable from o while o is not directly density-reachable from u because o is a core point while u is not a core point (it is border point).

Q5: DBSCAN Reachability and Clusters

Is n density-reachable (i.e. going only through core points) from e? Show the intermediate data points on the chain or the point where the chain breaks.



- n is density-reachable from e. For example valid path is e, c, f, j, n. There are more density-based paths. They may include points b, g in some combination with c, f, j, n.

Show the density-based clusters and the noise points

- For the points to be in the same cluster they must be density reachable from each other. Clusters are $C_1 = \{p, q, v\}$, $C_2 = \{r, s, t, k, w, x, h, d, a\}$ whre x, a are the border points for C_2, all other points are core points. Cluster $C_3 = \{m, i, e, b, c, f, g, j, n, o, u\}$ where m and u are border points. The only noise point is l.

Lecture Overview

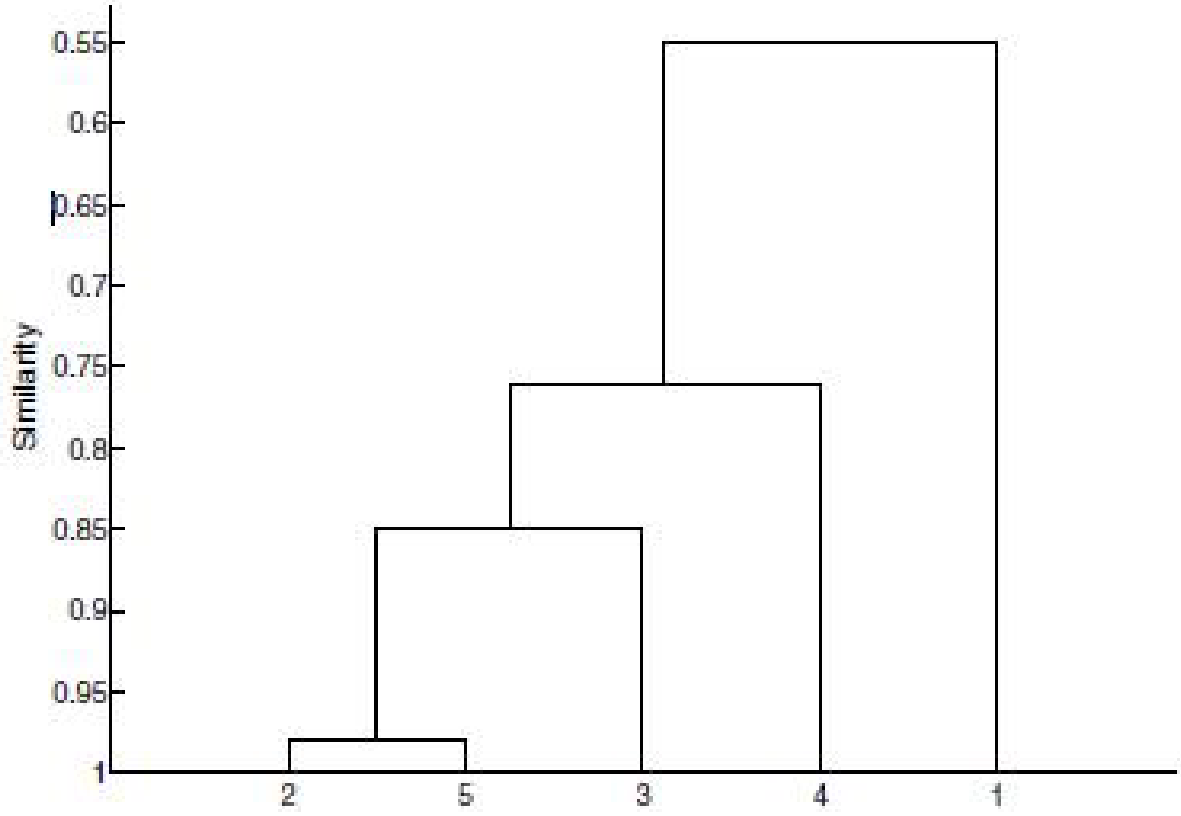
1. Final Structure
2. Naïve Bayes Classification
3. SVM Classifier
4. Hierarchical Clustering
5. K-means
6. DBSCAN
7. Cophenetic Correlation

Q6: Cophenetic Measures

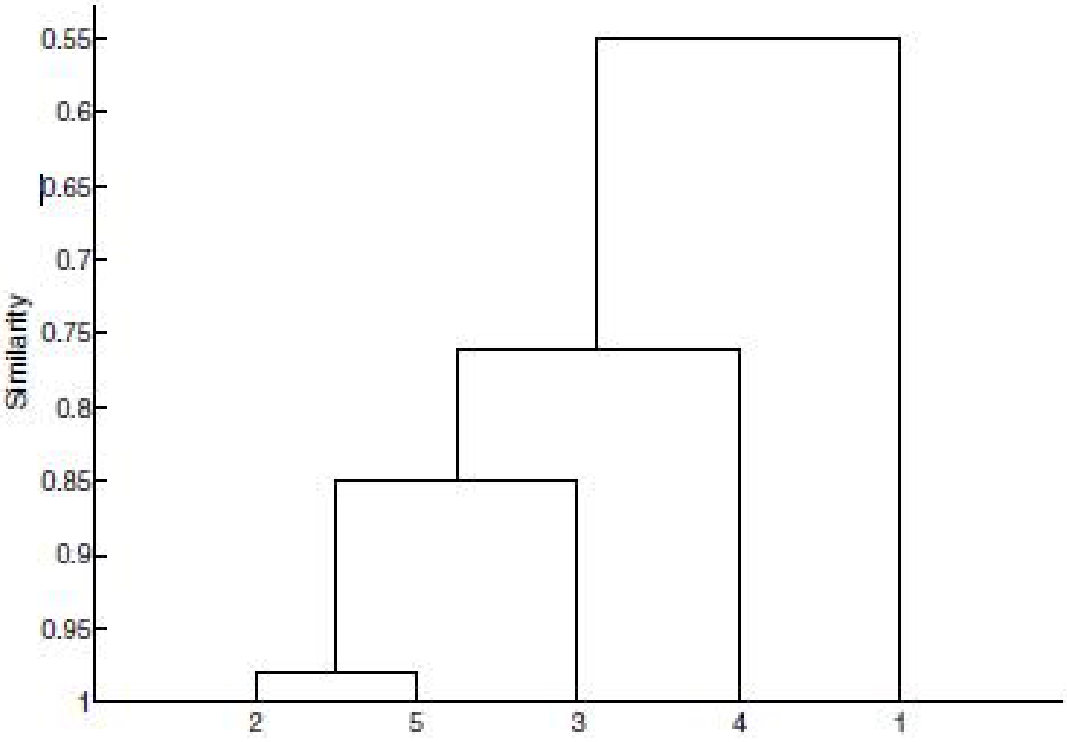
Compute the cophenetic correlation coefficient for the hierarchical clustering in Q3 single link.

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

Table 1



Q6: Cophenetic Similarities



Convert the dendrogram to cophenetic similarities matrix:

	P1	P2	P3	P4	P5
P1	1	0.55	0.55	0.55	0.55
P2		1	0.85	0.76	0.98
P3			1	0.76	0.85
P4				1	0.76
P5					1

Q6: Similarities to Dissimilarities to Vectors

Distances:

	P1	P2	P3	P4	P5
P1	1	0.1	0.41	0.55	0.35
P2	0.1	1	0.64	0.47	0.98
P3	0.41	0.64	1.0	0.44	0.85
P4	0.55	0.47	0.44	1.00	0.76
P5	0.35	0.98	0.85	0.76	1.00

	P1	P2	P3	P4	P5
P1	0	0.9	0.59	0.45	0.65
P2		0	0.36	0.53	0.02
P3			0	0.56	0.15
P4				0	0.24
P5					0

Cophenetic distances:

	P1	P2	P3	P4	P5
P1	1	0.55	0.55	0.55	0.55
P2		1	0.85	0.76	0.98
P3			1	0.76	0.85
P4				1	0.76
P5					1

	P1	P2	P3	P4	P5
P1	0	0.45	0.45	0.45	0.45
P2		0	0.15	0.24	0.02
P3			0	0.24	0.15
P4				0	0.24
P5					0

We treat both dissimilarities as vectors, using upper triangular part of original dissimilarity matrix not including the diagonal:

$$T = \begin{pmatrix} 0.9 \\ 0.59 \\ 0.45 \\ 0.65 \\ 0.36 \\ 0.53 \\ 0.02 \\ 0.56 \\ 0.15 \\ 0.54 \end{pmatrix} \text{ and } CT = \begin{pmatrix} 0.45 \\ 0.45 \\ 0.45 \\ 0.45 \\ 0.15 \\ 0.24 \\ 0.02 \\ 0.24 \\ 0.15 \\ 0.24 \end{pmatrix}$$

Q6 – Cophenetic Correlation Coefficient

- We treat both dissimilarities as vectors, using upper triangular part of original dissimilarity matrix not including the diagonal

$$T = \begin{pmatrix} 0.9 \\ 0.59 \\ 0.45 \\ 0.65 \\ 0.36 \\ 0.53 \\ 0.02 \\ 0.56 \\ 0.15 \\ 0.54 \end{pmatrix} \text{ and } CT = \begin{pmatrix} 0.45 \\ 0.45 \\ 0.45 \\ 0.45 \\ 0.15 \\ 0.24 \\ 0.02 \\ 0.24 \\ 0.15 \\ 0.24 \end{pmatrix}$$

- We compute Pearson correlation coefficient between T and CT as

$$corr(T, CT) = \frac{T_c \cdot CT_c}{\|T_c\| \|CT_c\|}$$

where T_c, CT_c are respective centered vectors so

$corr(T, CT) = 0.8039$