

제6강. 군집분석(1)

- 군집분석이란
- 거리, 유사성측도
- 계층적 군집분석
- R을 이용한 계층적 군집분석 사례

1. 군집분석이란

1) 군집분석(Cluster Analysis)이란?

- N개의 관찰치들을 대상으로 p개의 변수를 측정하였을 때,
 - 관측한 p개의 변수값을 이용하여
 - N개 관찰치들 사이의 유사성(similarity) 또는 비유사성(dissimilarity)의 정도를 측정하여
 - 서로 유사한 관찰치들을 하나의 군집으로 묶어주는 통계적 분석방법
- 특성
 - 같은 군집내의 관찰치들은 서로 유사하나,
 - 다른 군집에 속한 관찰치와는 상이함

1. 군집분석이란

2) 군집분석의 예

- 우리나라 도시들의 군집화
 - 우리나라 11개 대도시들 (서울, 부산, 대구, 광주, 대전, 인천, 수원, 청주, 춘천, 전주, 제주)을 대상
 - 면적, 인구, 소득수준, 취업률, 주택보급률 등과 같은 변수들을 관측한 후에
 - 이 변수들을 이용하여 각 도시들 사이의 유사성의 정도를 측정
 - 어느 도시들이 가장 유사한가?
 - 우리 나라 도시들은 몇 개의 군집으로 나누어 볼 수 있는가?
 - 정부정책의 의사결정에 군집화 정보가 활용될 수 있음

1. 군집분석이란

2) 군집분석의 예

- 상장회사들의 군집화
 - 증권시장에서 상장된 주식의 여러 가지 특징
 - 예를 들어 주식가격, 자본규모, 당기 순이익, 자기자본수익률 등을 측정한 후에
 - 상장회사들을 유사한 회사끼리 군집화
 - 경기전망에 따른 투자 전략에 군집화 정보가 활용 가능
- 환자들의 군집화
 - 환자들의 여러 가지 증상에 따라 환자들을 군집화
 - 군집별로 가장 효과적인 치료법 개발

1. 군집분석이란

2) 군집분석의 예

- 고객들의 군집화
 - 회사의 마케팅 담당부서는 고객에 관한 데이터를 수집
 - 고객이 요구하는 것을 파악함으로써 고객에 대한 서비스를 향상시키고자 노력
 - 수집하는 데이터는 고객의 연령, 교육수준, 소득수준, 결혼여부, 취업여부, 5세 이하 자녀의 수, 6세부터 13세까지의 자녀의 수, 14세 이상 자녀의 수 등과 같은 고객에 대한 실제 정보
 - 고객들을 몇 개의 소집단(군집)으로 구분지은 후 각 군집별로 독특한 마케팅 업무를 수행
 - 예) 여행선호 고객군집에 다양한 여행상품 제공
 - 각 군집의 독특한 정보를 가장 효율적으로 사용하여 최대의 광고효과

2. 유사성 측도

1) 유사성 측도의 정의

- 두 관찰치:

개 체 \ 변 수	1	2	3	...	p
	X_1	X_2	X_3	...	X_p
Y	Y_1	Y_2	Y_3	...	Y_p

- 유사성의 측정

1. 거리 (distance): 값이 작을수록 두 관찰치가 서로 유사한 것을 의미
 - ① 유클리디안 거리, 유클리디안 제곱거리, 세비체프(Chebychev) 거리, Block, City-Block, 또는 맨해튼(Manhattan)거리, 민코브스키(Minkowski) 거리, Customized 거리
2. 유사성 (similarity): 값이 클수록 두 관찰치가 서로 유사한 것을 의미
 - ① 코사인(Cosine) 값, 상관계수

2. 유사성 측도

1) 유사성 측도의 정의 - 거리중심

- 유클리디안(Euclidean) 거리: 두 관찰치사이의 거리측정

$$D(X, Y) = \sqrt{\sum_{i=1}^p (X_i - Y_i)^2}$$

- 유클리디안 제곱거리: 유클리디안 거리의 제곱값

$$D(X, Y) = \sum_{i=1}^p (X_i - Y_i)^2$$

- 세비체프(Chebyshev) 거리 (거리측정) : 두 관찰치 사이의 거리측정으로 변수값의 최대차이의 절대값

$$D(X, Y) = \max_i |X_i - Y_i|$$

2. 유사성 측도

1) 유사성 측도의 정의 - 거리중심

- Block, City-Block, 또는 맨해튼(Manhattan)거리
: 두 관찰치 사이의 거리로 각 변수값 차이의 절대값의 합

$$D(X, Y) = \sum_{i=1}^p |X_i - Y_i|$$

- 민코브스키(Minkowski) 거리: p는 양의 정수값으로 디폴트값은 2

$$D(X, Y) = \left(\sum_{i=1}^p |X_i - Y_i|^p \right)^{\frac{1}{p}}$$

- Customized 거리: p와 r은 양의 정수값으로 디폴트값은 2

$$D(X, Y) = \left(\sum_{i=1}^p |X_i - Y_i|^p \right)^{\frac{1}{r}}$$

2. 유사성 측도

1) 유사성 측도의 정의 - 유사성중심

- 코사인(Cosine) 값: 두 벡터 X와 Y사이의 코사인 값

$$S(X, Y) = \frac{\sum_{i=1}^p X_i Y_i}{\sqrt{\sum_{i=1}^p X_i^2 \cdot \sum_{i=1}^p Y_i^2}}$$

- 상관계수: X와 Y 사이의 상관계수

$$S(X, Y) = \frac{\sum_{i=1}^p (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^p (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^p (Y_i - \bar{Y})^2}}$$

- 가장 일반적으로 이용되는 방법은 유클리디안 거리 또는 상관계수

2. 유사성 측도

1) 유사성 측도의 정의

- 주의점: 거리중심 유사성 측도는 각 변수들의 측정단위에 영향을 받음
- 측정단위의 영향력을 없애기 위하여 군집분석에서는 각 변수들을 표준화하여 사용함

$$Z_i = \frac{W_i - \bar{W}}{S_W}, \quad i = 1, 2, \dots, N$$

▪ 여기서

$$S_W = \sqrt{\frac{\sum (W_i - \bar{W})^2}{N - 1}}$$

2. 유사성 측도

1) 유사성 측도의 정의

- 유사성행렬(similarity matrix)

: 모든 관찰치들 쌍에 대하여 유사성을 측정하여 정리해 놓은 행렬

	1	2	3	4	5
1	1	0.584	0.615	0.601	0.570
2		1	0.576	0.530	0.526
3			1	0.940	0.875
4				1	0.877
5					1

2. 유사성 측도

2) 유사성 측도의 예제

- 20개 맥주브랜드에 대해서 칼로리량, 염분량, 알코올농도, 가격등에 대하여 조사한 내용

변수명	의미
BEER	맥주이름
X1: CALORIES	12온스당 칼로리량
X2: SODIUM	12온스당 염분량(mg)
X3: ALCOHOL	알코올농도(%)
X4: COST	12온스당 가격(\$)

2. 유사성 측도

2) 유사성 측도의 예제

ID	BEER	CALORIES	SODIUM	ALCOHOL	COST
1	BUDWEISER	144	15	4.7	0.43
2	SCHLITZ	151	19	4.9	0.43
3	LOWENBRAU	157	15	4.9	0.48
4	KRONENBOURG	170	7	5.2	0.73
5	HEINEKEN	152	11	5.0	0.77
6	OLD MILWAUKEE	145	23	4.6	0.28
7	AUGSBURGER	175	24	5.5	0.40
8	STROHS BOHEMIAN STYLE	149	27	4.7	0.42
9	MILLER LITE	99	10	4.3	0.43
10	BUDWEISER LIGHT	113	8	3.7	0.44
11	COORS	140	18	4.6	0.44
12	COORS LIGHT	102	15	4.1	0.46
13	MICHELOB LIGHT	135	11	4.2	0.50
14	BECKS	150	19	4.7	0.76
15	KIRIN	149	6	5.0	0.79
16	PABST EXTRA LIGHT	68	15	2.3	0.38
17	HAMMS	136	19	4.4	0.43
18	HEILEMANS OLD STYLE	144	24	4.9	0.43
19	OLYMPIA GOLD LIGHT	72	6	2.9	0.46
20	SCHLITZ LIGHT	97	7	4.2	0.47

2. 유사성 측도

2) 유사성 측도의 예제

- 두 맥주 BUDWEISER와 SCHLITZ를 대상으로 거리 측정

BEER	X_1	X_2	X_3	X_4
BUDWEISER	144	15	4.7	0.43
SCHLITZ	151	19	4.9	0.43

- 각 변수들의 측정단위가 다르기 때문에 표준화 필요함

	X_1	X_2	X_3	X_4
평균(\bar{X})	132.4	14.95	4.44	0.50
표준편차(S)	30.26	6.58	0.76	0.14

2. 유사성 측도

2) 유사성 측도의 예제

- BUDWEISER에서 변수 X1 (CALORIES)의 표준화한 값

$$Z_1 = \frac{X_1 - \bar{X}}{S} = \frac{144 - 132.4}{30.2575} = 0.3834$$

- 같은 방식으로

BUDWEISER와 SCHLITZ 맥주의 표준화한 값

BEER	Z_1	Z_2	Z_3	Z_4
BUDWEISER	0.3834	0.0076	0.3422	-0.4626
SCHLITZ	0.6147	0.6154	0.6054	-0.4626

2. 유사성 측도

2) 유사성 측도의 예제 - 거리 중심

(1) 유클리디안(Euclidean) 거리 :

$$\begin{aligned} D_1 &= \sqrt{\sum_{z=1}^4 (Z_{z1} - Z_{z2})^2} = [(0.3833 - 0.6147)^2 + (0.0076 - 0.6155)^2 \\ &\quad (0.3421 - 0.6053)^2 + (-0.5 - (-0.5))^2]^{\frac{1}{2}} \\ &= \sqrt{0.4922} \\ &= 0.7016 \end{aligned}$$

(2) 유클리디안(Euclidean) 제곱거리 :

$$D_2 = \sum_{z=1}^4 (Z_{z1} - Z_{z2})^2 = 0.4922$$

2. 유사성 측도

2) 유사성 측도의 예제 - 유사성 중심

(3) 코사인(Cosine) 값 :

$$S_1 = \frac{\sum_{i=1}^4 Z_{1i} \cdot Z_{2i}}{\sqrt{\sum_{i=1}^4 Z_{1i}^2 \cdot \sum_{i=1}^4 Z_{2i}^2}} = \frac{0.6616}{\sqrt{0.4782 \times 1.3371}} = 0.8274$$

(4) 상관계수 :

$$r = \frac{S_{12}}{\sqrt{S_1^2} \sqrt{S_2^2}} = \frac{0.1895}{\sqrt{0.1533} \sqrt{0.2886}} = 0.9012$$

군집분석방법

1. 계층적 군집분석(hierarchical cluster)
2. 비계층적 군집분석(nonhierarchical cluster)

3. 계층적 군집분석

전체 관찰치는 개별 군집으로 취급한다. 이러한 개별 관찰치를 C_1, C_2, \dots, C_n 이라 했을 때, i 번째 관찰치와 j 번째 관찰치 사이의 거리측도를 d_{ij} 라 하자.


	1	2	3	4	5
1	0	5.1	3	4	6.5
2		0	6	7	8.1
3			0	0.6	2
4				0	1.1
5					0

3. 계층적 군집분석

1) 계층적 군집분석 절차 : 최단연결법

1단계 : 가장 적은 거리측도 (혹은 높은 유사성 측도) 값을 가지는 두 개의 군집을 찾는다.

	1	2	3	4	5
1	0	5.1	3	4	6.5
2		0	6	7	8.1
3			0	0.6	2
4				0	1.1
5					0



	1	2	(3, 4)	5
1	0	5.1	?	6.5
2		0	?	8.1
(3, 4)			0	?
5				0

관찰치3과 관찰치4의 거리가 0.6 으로 가장 작으므로 관찰치 3과 관찰치 4를 제일 먼저 연결하여 (3, 4)를 하나의 개체로 취급.
그러면 묶임의 대상이 되는 관찰치는 (1), (2), (5)와 (3, 4) 4개가 됨.

3. 계층적 군집분석

1) 계층적 군집분석 절차 : 최단연결법/최장연결법

1단계 : 가장 적은 거리측도 (혹은 높은 유사성 측도) 값을 가지는 두 개의 군집을 찾는다.

	1	2	3	4	5
1	0	5.1	3	4	6.5
2		0	6	7	8.1
3			0	0.6	2
4				0	1.1
5					0

→

	1	2	(3, 4)	5
1	0	5.1	3	6.5
2		0	6	8.1
(3, 4)			0	1.1
5				0

〈최단연결법〉

	1	2	(3, 4)	5
1	0	5.1	4	6.5
2		0	7	8.1
(3, 4)			0	2
5				0

〈최장연결법〉

관찰치 1, 2, 5와 관찰치(3, 4)와의 거리측정에서 최단(최장)연결법은 거리의 가장 작은(큰) 값을 이용한다. 즉 관찰치 1과 3의 거리는 3이고, 관찰치 1과 4의 거리는 4이므로 관찰치 1과 (3, 4)의 거리는 3(4)이다.

3. 계층적 군집분석

1) 계층적 군집분석 절차 : 최단연결법/최장연결법

2단계 : (3,4)와 5의 거리가 가장 작으므로 (3,4,5)를 연결



관찰치 1과 관찰치(3, 4, 5)와의 거리는 (1)과 (3,4)의 거리가 3, (1)과 (5)와의 거리가 6.5 이므로 최단(최장)연결법 정의에서 거리의 가장 작은(큰) 값 3(6.5) 이 됨.

3. 계층적 군집분석

1) 계층적 군집분석 절차 : 최단연결법/최장연결법

〈최단연결법〉 3단계 : (3,4,5)와 1의 거리가 가장 작으므로 (1,3,4,5)를 연결

$$\begin{array}{c}
 \begin{array}{c} 1 \quad 2 \quad (3, 4, 5) \\ \begin{bmatrix} 0 & 5.1 & 3 \\ & 0 & 6 \\ (3, 4, 5) & & 0 \end{bmatrix} \end{array}
 \end{array}
 \rightarrow
 \begin{array}{c}
 \begin{array}{c} 2 \quad (1, 3, 4, 5) \\ \begin{bmatrix} 0 & 5.1 \\ (1, 3, 4, 5) & 0 \end{bmatrix} \end{array}
 \end{array}$$

4단계 : 관찰치 (2)와 (1,3,4,5)를 연결. 거리는 5.1 이 됨.

〈최장연결법〉 3단계 : (1)과 (2)의 거리가 가장 작으므로 (1,2)를 연결

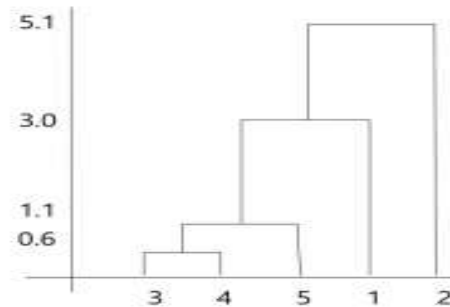
$$\begin{array}{c}
 \begin{array}{c} 1 \quad 2 \quad (3, 4, 5) \\ \begin{bmatrix} 0 & 5.1 & 6.5 \\ & 0 & 8.1 \\ (3, 4, 5) & & 0 \end{bmatrix} \end{array}
 \end{array}
 \rightarrow
 \begin{array}{c}
 \begin{array}{c} (1, 2) \quad (3, 4, 5) \\ \begin{bmatrix} 0 & 8.1 \\ (3, 4, 5) & 0 \end{bmatrix} \end{array}
 \end{array}$$

4단계 : 관찰치 (1,2)와 (3,4,5)를 연결. 거리는 8.1 이 됨.

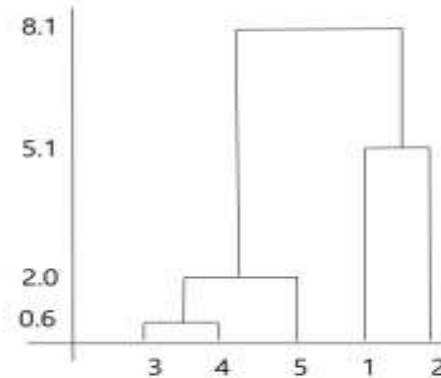
3. 계층적 군집분석

1) 계층적 군집분석 절차 : 덴드로그램

〈최단연결법〉



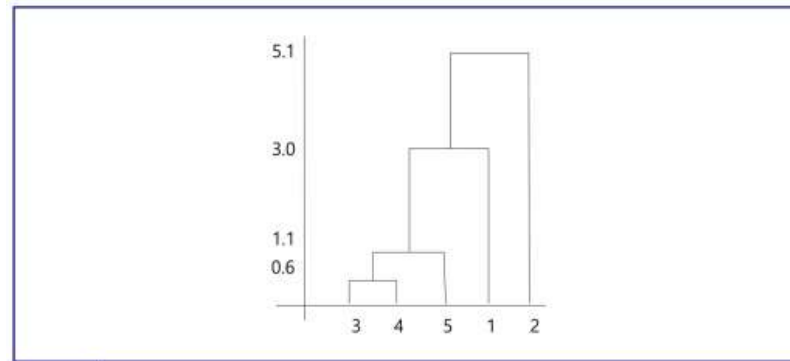
〈최장연결법〉



3. 계층적 군집분석

덴드로그램 (Dendrogram)

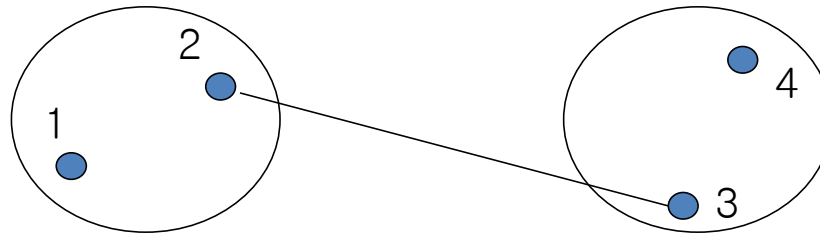
- 계층적 군집분석으로 관찰치들을 군집화한 후에 연결결과를 순서에 따라 정리해 놓은 그림
- 덴드로그램에 의하여 계층적 군집분석 방법의 결과를 이해할 수 있음
- 가지의 높이는 관찰치와 군집이 통합될 때의 거리측도(혹은 유사성 측도)와 비례



3. 계층적 군집분석

2) 계층적 군집분석 연결방법

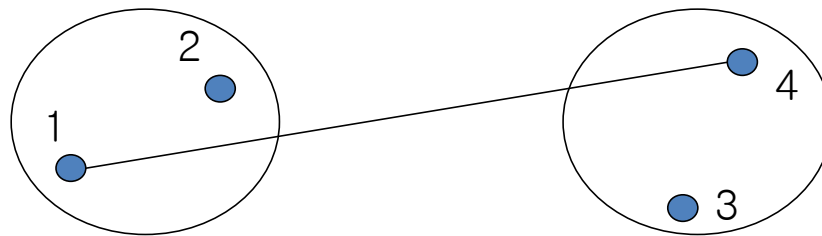
(1) 최단 연결방법(Nearest-neighbor) : 각 관찰치 또는 군집 사이의 거리가 가장 작은 값 (유사성이 가장 큰 값)을 기준으로 연결해가는 방법.
이 방법은 Single linkage 방법이라 불리기도 함



3. 계층적 군집분석

2) 계층적 군집분석 연결방법

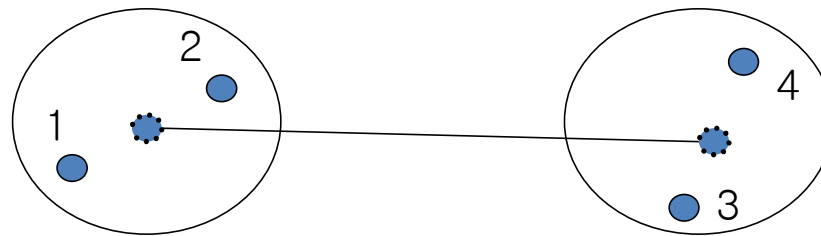
(2) 최장연결법(Furthest-neighbor) : 각 관찰치 또는 군집사이의 거리가 가장 큰 값 (유사성이 가장 작은 값)을 기준으로 연결해가는 방법.
이 방법은 Complete linkage 방법이라 불리기도 함



3. 계층적 군집분석

2) 계층적 군집분석 연결방법

(3) 중심연결법(Centroid linkage) : 각 군집내의 거리(유사성)를 군집의 중심점을 기준으로 연결해 가는 방법

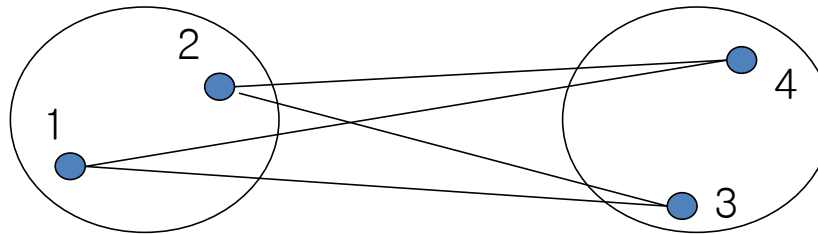


3. 계층적 군집분석

2) 계층적 군집분석 연결방법

(4) 평균연결법(Average linkage) : 두 군집의 관찰치간 모든 거리들의 평균을 군집간 거리로 측정하여 연결해 가는 방법.

아래 그림에서 두 군집간 거리는 $\frac{(d_{13} + d_{23} + d_{14} + d_{24})}{4}$ 으로 정의된다.



3. 계층적 군집분석

2) 계층적 군집분석 연결방법

(5) 와드의 방법(Ward's Method) 연결 가능한 군집조합 중 연결된 후에 군집내 제곱합을 계산하여 최소 제곱합을 가지게 되는 군집끼리 연결시키는 방법

먼저 ESS_j 를 j 번째 군집의 제곱합이라 하자.

여기서 제곱합은 군집중심으로부터 관찰치까지의 거리의 제곱합을 의미한다.

만약 현 단계에서 J 개의 군집이 있다면 총 제곱합은 $ESS = \sum_{j=1}^J ESS_j$ 이다.

임의의 두 개 군집을 통합하고 다시 ESS 를 계산할 수 있게 되는데,
이때 ESS 가 최소가 되는 두 개 군집을 연결하는 방법

3. 계층적 군집분석

4) 기타 고려사항

(1) 방법의 선택

- 최단연결법은 하나의 큰 군집화를 순차적으로 만들어가는 경향이 있는데 비하여 최장연결법은 비슷한 크기로 여러 개 군집을 만들어서 서로 연결시키는 경향
- 기타 방법들은 최단연결법의 특징과 최장연결법의 특징의 중간에 놓여있음
- 여러 군집분석 방법을 수행하여 보고 비교해보는 것도 좋은 접근방법
- 군집분석 방법의 결과들이 매우 유사하다면, 연결방법에 관한 선택은 심각한 문제가 되지 않음

3. 계층적 군집분석

4) 기타 고려사항

(2) 군집분석 결과의 검증

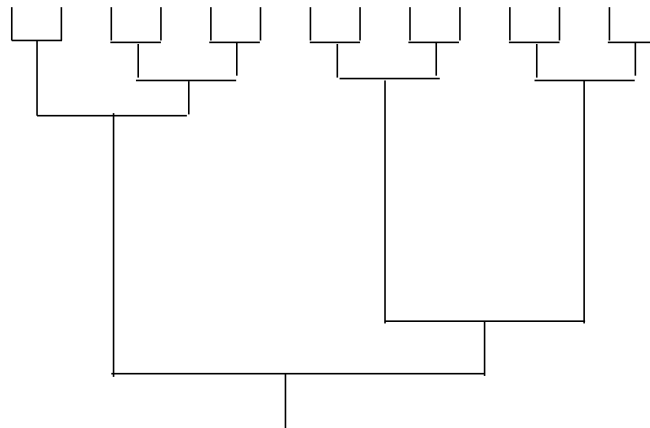
- 묶인 군집들이 상식적으로 타당한 군집인지 다변량 그림 등을 그려서 확인해 보는 방법
 - ① 변수가 두 개인 경우에 산점도,
 - ② 주성분 분석에서 구해진 상위 2개 주성분을 이용한 산점도,
 - ③ 변수가 3개인 경우에 3차원 산점도 혹은 bubble 그림,
 - ④ 주성분 분석에서 구해진 상위 3개 주성분을 이용한 3차원 산점도 및 bubble 그림,
 - ⑤ star 그림과 같은 다변량 그림 (andrews 그림과 Chernoff 얼굴그림도 가능)

3. 계층적 군집분석

4) 기타 고려사항

(3) 군집의 수 결정

- ❖ 수형도에서 거리측도의 값이 큰 변화를 보이는 위치에서 군집의 수를 결정하는 방법. 3개 군집의 예



- ❖ 와드방법을 사용한 계층적 군집분석인 경우에 ESS 값이 큰 변화를 보이는 위치에서 군집의 수를 결정하는 방법

4. R 군집분석 사례분석 1

1) 데이터 설명

- ❖ 20개 맥주상표를 대상으로 가격, 칼로리, 염분, 알코올농도 등을 측정한 자료
- ❖ 각 변수의 관측단위가 다르기 때문에 각 변수들을 표준화하여 이용하는 것이 바람직함

변수명	의미
BEER X1: CALORIES X2: SODIUM X3: ALCOHOL X4: COST	맥주이름 12온스당 칼로리량 12온스당 염분량(mg) 알코올농도(%) 12온스당 가격(\$)

4. R 군집분석 사례분석 1

1) 데이터 읽기 beerbrand.csv



	calories,	sodium,	alcohol,	cost
BUDWEISER,	144,	15,	4.7,	0.43
SCHLITZ,	151,	19,	4.9,	0.43
LOWENBRAU,	157,	15,	4.9,	0.48
KRONENBOURG,	170,	7,	5.2,	0.73
HEINEKEN,	152,	11,	5.0,	0.77
OLD MILWAUKEE,	145,	23,	4.6,	0.28
AUGSBERGER,	175,	24,	5.5,	0.40
STROHS BOHEMIAN STYLE,	149,	27,	4.7,	0.42
MILLER LITE,	99,	10,	4.3,	0.43
BUDWEISER LIGHT,	113,	8,	3.7,	0.44
COORS,	140,	18,	4.6,	0.44
COORS LIGHT,	102,	15,	4.1,	0.46
MICHELOB LIGHT,	135,	11,	4.2,	0.50
BECKS,	150,	19,	4.7,	0.76
KIRIN,	149,	6,	5.0,	0.79
PABST EXTRA LIGHT,	68,	15,	2.3,	0.38
HAMMS,	136,	19,	4.4,	0.43
HEILEMANS OLD STYLE,	144,	24,	4.9,	0.43
OLYMPIA GOLD LIGHT,	72,	6,	2.9,	0.46
SCHLITZ LIGHT,	97,	7,	4.2,	0.47

4. R 군집분석 사례분석 1

데이터 읽기

```
> beer = read.csv("c:/data/mva/beerbrand.csv", header=T, row.names=1)  
> head(beer)
```

	calories	sodium	alcohol	cost
BUDWEISER	144	15	4.7	0.43
SCHLITZ	151	19	4.9	0.43
LOWENBRAU	157	15	4.9	0.48
KRONENBOURG	170	7	5.2	0.73
HEINEKEN	152	11	5.0	0.77
OLD MILWAUKEE	145	23	4.6	0.28

```
> summary(beer)
```

calories		sodium		alcohol		cost	
Min.	: 68.0	Min.	: 6.00	Min.	:2.30	Min.	:0.2800
1st Qu.:	110.2	1st Qu.:	9.50	1st Qu.:	4.20	1st Qu.:	0.4300
Median	:144.0	Median	:15.00	Median	:4.65	Median	:0.4400
Mean	:132.4	Mean	:14.95	Mean	:4.44	Mean	:0.4965
3rd Qu.:	150.2	3rd Qu.:	19.00	3rd Qu.:	4.90	3rd Qu.:	0.4850
Max.	:175.0	Max.	:27.00	Max.	:5.50	Max.	:0.7900

4. R 군집분석 사례분석 1

2) 자료 표준화

```
> zbeer = scale(beer)
> round(apply(zbeer, 2, mean), 3)
calories    sodium    alcohol      cost
          0          0          0          0
> round(apply(zbeer, 2, sd), 3)
calories    sodium    alcohol      cost
          1          1          1          1
>
```

```
> # 0-1 변환
> library(caret)
> z01_beer = preProcess(beer,
  method='range')
> z01_model = preProcess(beer,
  method='range')
> z01_beer = predict(z01_model, beer)
> summary(z01_beer)
```

```
> # 0-1 변환 (2)
> maxX = apply(beer, 2, max)
> minX = apply(beer, 2, min)
> z01X = scale(beer, center=minX,
  scale=maxX-minX)
> summary(z01X)
```

4. R 군집분석 사례분석 1

3) 거리행렬 계산하기

```
> zbeer_euc = dist(zbeer)
> zbeer_man = dist(zbeer, "manhattan")

> zbeer_euc[1]
[1] 0.7015818
> zbeer_man[1]
[1] 1.102367
```

4. R 군집분석 사례분석 1

4) 계층적 군집분석 - 최단연결법

```
> hc_s = hclust(zbeer_euc, method='single')  
> hc_s
```

Call:

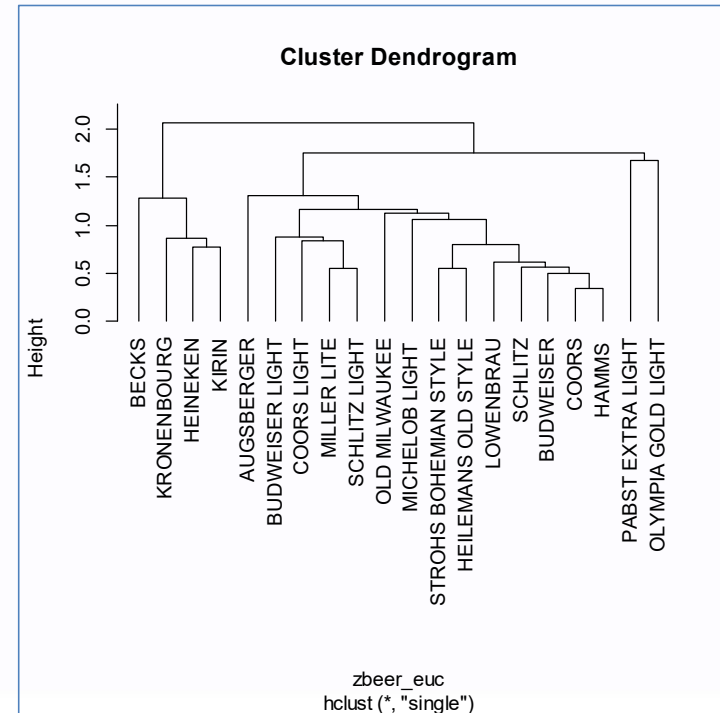
```
hclust(d = zbeer_euc, method = "single")
```

Cluster method : single

Distance : euclidean

Number of objects: 20

```
> plot(hc_s, hang=-1)
```



4. R 군집분석 사례분석 1

5) 계층적 군집분석 - 최장연결법

```
> hc_c = hclust(zbeer_euc, method='complete')  
> hc_c
```

Call:

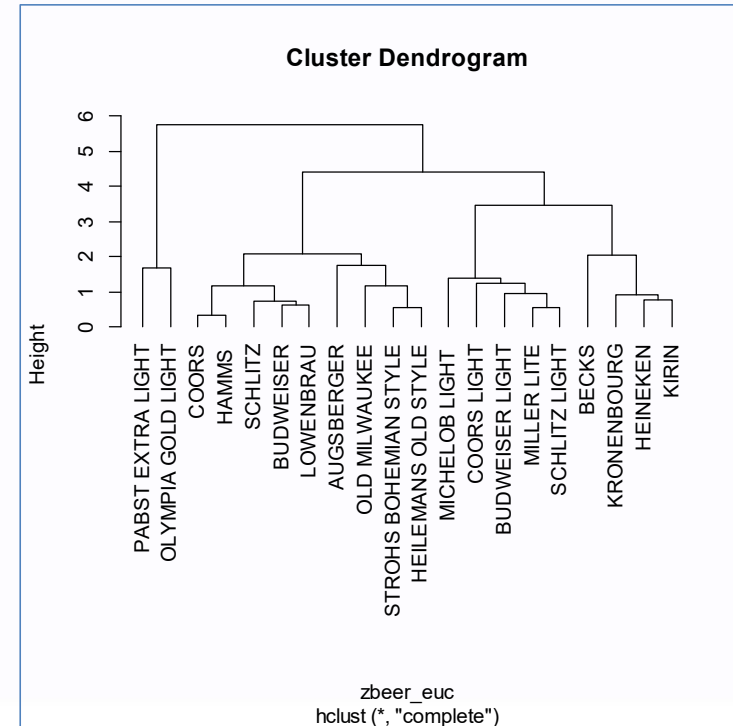
```
hclust(d = zbeer_euc, method = "complete")
```

Cluster method : complete

Distance : euclidean

Number of objects: 20

```
> plot(hc_s, hang=-1)
```



4. R 군집분석 사례분석 1

6) 계층적 군집분석 - 중심연결법

```
> hc_cen = hclust(zbeer_euc, method="centroid")  
> hc_cen
```

Call:

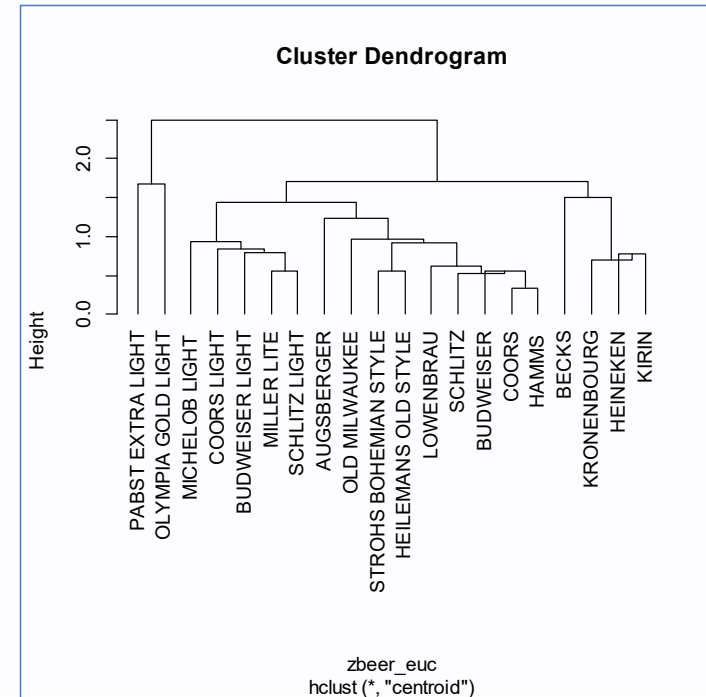
```
hclust(d = zbeer_euc, method = "centroid")
```

Cluster method : centroid

Distance : euclidean

Number of objects: 20

```
> plot(hc_cen, hang=-1)
```



4. R 군집분석 사례분석 1

7) 계층적 군집분석 - 와드의 방법

```
> hc_w = hclust(zbeer_euc, method="ward.D")  
> hc_w
```

Call:

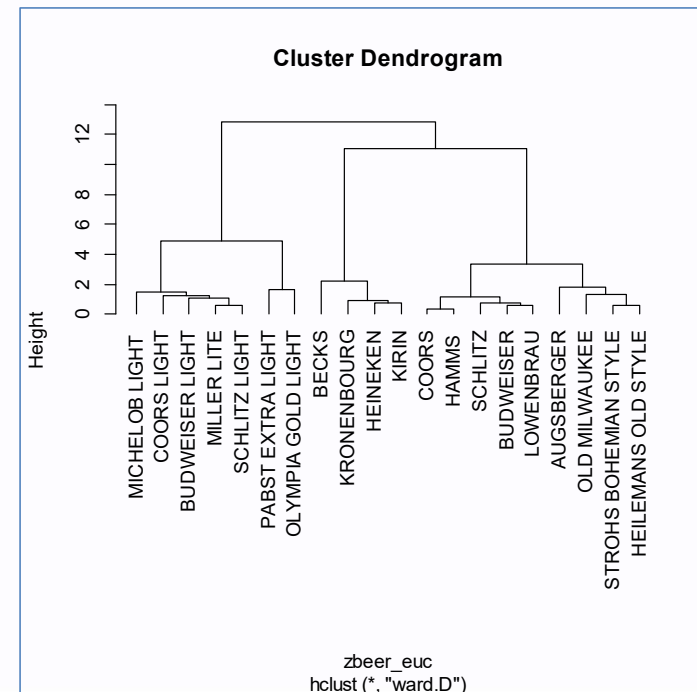
```
hclust(d = zbeer_euc, method = "ward.D")
```

Cluster method : ward.D

Distance : euclidean

Number of objects: 20

```
> plot(hc_w, hang=-1)
```



4. R 군집분석 사례분석 1

8) 소속 군집 알기

```
> hc_cen24 = cutree(hc_cen, 2:4)
```

```
> hc_cen24
```

	2	3	4
BUDWEISER	1	1	1
SCHLITZ	1	1	1
LOWENBRAU	1	1	1
KRONENBOURG	1	2	2
HEINEKEN	1	2	2
OLD MILWAUKEE	1	1	1
AUGSBERGER	1	1	1
STROHS BOHEMIAN STYLE	1	1	1
MILLER LITE	1	1	1
BUDWEISER LIGHT	1	1	1
COORS	1	1	1
COORS LIGHT	1	1	1
MICHELOB LIGHT	1	1	1
BECKS	1	2	2
KIRIN	1	2	2
PABST EXTRA LIGHT	2	3	3
HAMMS	1	1	1
HEILEMANS OLD STYLE	1	1	1
OLYMPIA GOLD LIGHT	2	3	4
SCHLITZ LIGHT	1	1	1

```
>
```

다음시간에는

7강 군집분석 (2)

 수고했습니다.