

비지도 분류모형 - 클러스터링 & 잠재계층분석



Fall, 2021

Syllabus

Week	Date	Topic	Note
1	9/6(월)	R Basic - R 기초 문법 학습	
2	9/13(월)	R Basic – Data Manipulation I	과제#1
3	9/20(월) (추석)	<추석> (보충영상) R Basic - Data Manipulation II	
4	9/27(월)	Descriptive Analytics I - 데이터 요약하기/상관관계/차이검증	과제#2
5	10/4(월) (대체공휴일)	<대체공휴일> (보충영상) Descriptive Analytics II - 데이터 시각화	과제#2
6	10/11(월) (대체공휴일)	<대체공휴일> (보충영상) Supplementary Topic I - 외부 데이터 수집 (정적 콘텐츠 수집)	과제#4 과제#3
7	10/18(월)	Predictive Analytics I – Linear regression	
8	10/25(월)	Predictive Analytics II – Logistic Regression	시험 대체 수업
9	11/1(월)	Predictive Analytics III – Clustering & Latent Class Analysis	과제#4
10	11/8(월)	Predictive Analytics IV – Tree-based Model and Bagging (Random Forest)	
11	11/15(월)	Predictive Analytics V – Association Rules	
12	11/22(월)	Prescriptive Analytics I – Linear Programming	과제#5
13	11/29(월)	Prescriptive Analytics II – Data Envelopment Analysis (DEA)	
14	12/6(월)	Prescriptive Analytics III – Integer Programming	과제#6
15	12/13(월)	Prescriptive Analytics IV – Simulation	Quiz
16	12/20(월)	Final Presentation	

Contents

	군집분석 (Cluster Analysis)	잠재계층분석 (Latent Class Analysis : LCA)
공통점	<ul style="list-style-type: none">▪ 비지도 학습▪ Y변수(Label)이 따로 없음▪ 집단의 수(k)를 직접 정해줘야 함	
집단 구분 변수	<ul style="list-style-type: none">▪ 수치형 변수▪ 연속형 변수	<ul style="list-style-type: none">▪ 범주형 변수▪ 이산형 변수
집단의 수 결정 방법	<ul style="list-style-type: none">▪ 군집 내 개체 간 거리의 제곱합	<ul style="list-style-type: none">▪ 정보통계량(AIC, BIC)

Lecture 9-1

데이터 간 거리는
어떻게 측정하는가?

데이터 간 거리를 어떻게 측정할 수 있을까?

데이터가 서로 얼마나 유사한지 혹은 유사하지 않은지를 기준으로 하며, 대표적으로 데이터 간 거리(비유사성)는 유클리드 거리, 유사성은 코사인 유사도를 측도로 사용함

유클리드 거리 (Euclidean Distance)

- 두 개체 간 유클리디안 공간 상의 거리(Distance)를 의미함
- 우측 formula와 같이 계산되는 거리를 '유클리드 거리'라고 정의함

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + \dots + (x_{ip} - x_{jp})^2}$$
$$= \sum_{k=1}^p (x_{ik} - x_{jk})^2$$

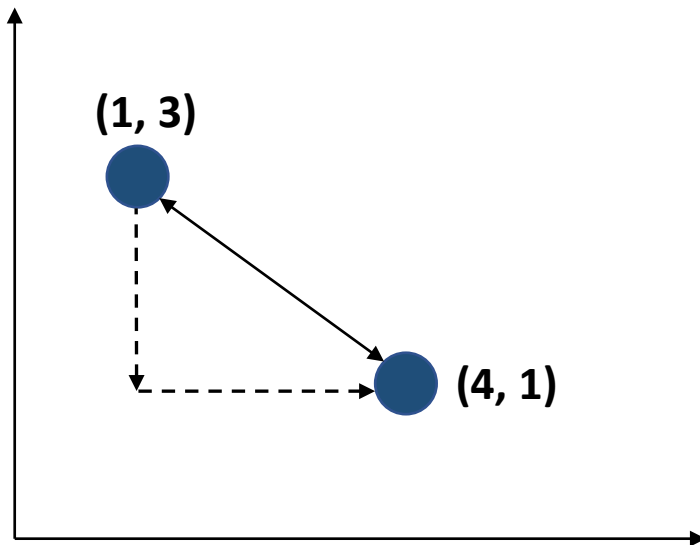
코사인 유사도 (Cosine Similarity)

- 거리(Distance)는 비유사도를 기준으로 하지만 때로는 유사도(Similarity)에 기반한 군집분석이 더 적합한 경우도 있음
- 유사도는 -1에서 1사이의 값으로만 주어져 상관관계수 처럼 해석할 수 있음

$$\cos(\theta)$$
$$= \frac{\sum_{k=1}^p x_{ik} * x_{jk}}{\sqrt{\sum_{k=1}^p (x_{ik})^2} \sqrt{\sum_{k=1}^p (x_{jk})^2}}$$

데이터 간 거리를 어떻게 측정할 수 있을까?

예 : 유클리드 거리(Euclidean distance)와 코사인 유사도(Cosine Similarity)



유클리드 거리(Euclidean distance)

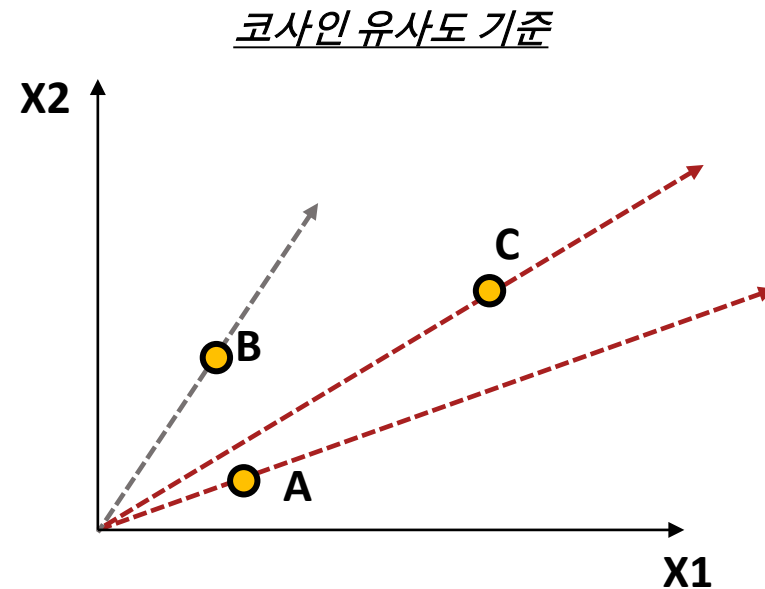
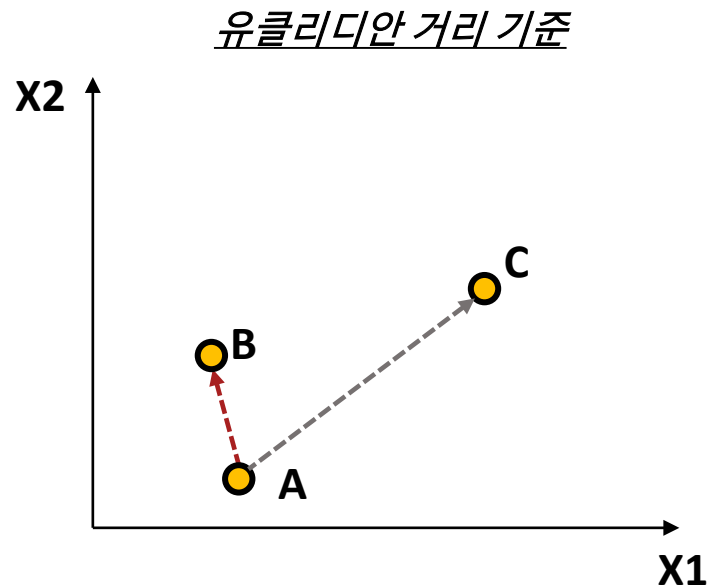
$$\sqrt{(4 - 1)^2 + (3 - 1)^2} = \sqrt{13} \approx 3.61$$

코사인 유사도(Cosine similarity)

$$\frac{(4 * 1) + (1 * 3)}{\sqrt{4^2 + 1^2} \sqrt{1^2 + 3^2}} \approx 0.537$$

※ 참고 – 유클리디언 거리와 코사인 유사도의 차이

A는 B와 유사한가, C와 유사한가?



“거리”는 벡터 공간 내 실질적인 거리의 물리적 “크기”를 측정하지만,
유사도는 “방향”만 고려한다.

데이터 간 거리를 어떻게 측정할 수 있을까?

3차원 이상의 데이터의 거리는 어떻게 구할 수 있을까?

구분	X1	X2	X3	X4	...
A소비자	4.2	3.3	2.2	12.2	...
B소비자	6.2	3.6	3.6	6.6	...
C소비자	1.7	2.2	7.7	17.4	...
⋮	⋮	⋮	⋮	⋮	⋮

(X1, X2) 이용(2차원)한 A와 B의 거리 = $\sqrt{(4.2 - 6.2)^2 + (3.3 - 3.6)^2} \approx \mathbf{2.02}$

(X1, X2, X3) 이용(3차원)한 A와 B의 거리 = $\sqrt{(4.2 - 6.2)^2 + (3.3 - 3.6)^2 + (2.2 - 3.6)^2} \approx \mathbf{2.459}$

(X1, X2, X3, X4) 이용(4차원)한 A와 B의 거리 = $\sqrt{(4.2 - 6.2)^2 + (3.3 - 3.6)^2 + (2.2 - 3.6)^2 + (12.2 - 6.6)^2} \approx \mathbf{6.11}$

데이터 간 거리를 어떻게 측정할 수 있을까?

3차원 이상의 데이터의 거리는 어떻게 구할 수 있을까?

구분	X1	X2	X3	X4	...
A소비자	4.2	3.3	2.2	12.2	...
B소비자	6.2	3.6	3.6	6.6	...
C소비자	1.7	2.2	7.7	17.4	...
⋮	⋮	⋮	⋮	⋮	⋮

$$(X1, X2) \text{ 이용(2차원)한 A와 B의 유사도} = \frac{(4.2 * 6.2) + (3.3 * 3.6)}{\sqrt{(4.2)^2 + (3.3)^2} \sqrt{(6.2)^2 + (3.6)^2}} \doteq \mathbf{0.99}$$

$$(X1, X2, X3) \text{ 이용(3차원)한 A와 B의 유사도} = \frac{(4.2 * 6.2) + (3.3 * 3.6) + (2.2 * 3.6)}{\sqrt{(4.2)^2 + (3.3)^2 + (2.2)^2} \sqrt{(6.2)^2 + (3.6)^2 + (3.6)^2}} \doteq \mathbf{0.989}$$

$$(X1, X2, X3, X4) \text{ 이용(4차원)한 A와 B의 유사도} = \frac{(4.2 * 6.2) + (3.3 * 3.6) + (2.2 * 3.6) + (12.2 * 6.6)}{\sqrt{(4.2)^2 + (3.3)^2 + (2.2)^2 + (12.2)^2} \sqrt{(6.2)^2 + (3.6)^2 + (3.6)^2 + (6.6)^2}} \doteq \mathbf{0.90}$$

자료의 정규화(Standardization)

만약, 다음과 같은 경우 어떤 문제가 발생하겠는가?

구분	외식 지출액(원)	나이 (세)	월 평균 외식횟수(회)	...
A소비자	300,000	22	5.3	...
B소비자	500,000	32	15.2	...
C소비자	1,200,000	29	10.2	...
⋮	⋮	⋮	⋮	⋮

A와 B의 유클리드 거리

$$D_{A,B} = \sqrt{(300,000 - 500,000)^2 + (22 - 32)^2 + (5.3 - 15.2)^2}$$

정규화되지 않은 자료를 바탕으로 거리를 구할 경우,
Scale 이 큰 변수에 의해 거리가 결정되는 경향이 큼

자료의 정규화(Standardization)

각 변수를 열 단위로 정규화(Standardization)

구분	외식 지출액(원)	나이 (세)	월 평균 외식횟수(회)	...
A소비자	0.5	0.7	1.0	...
B소비자	0.6	0.5	0.3	...
C소비자	2.0	0.3	0.7	...
⋮	⋮	⋮	⋮	⋮

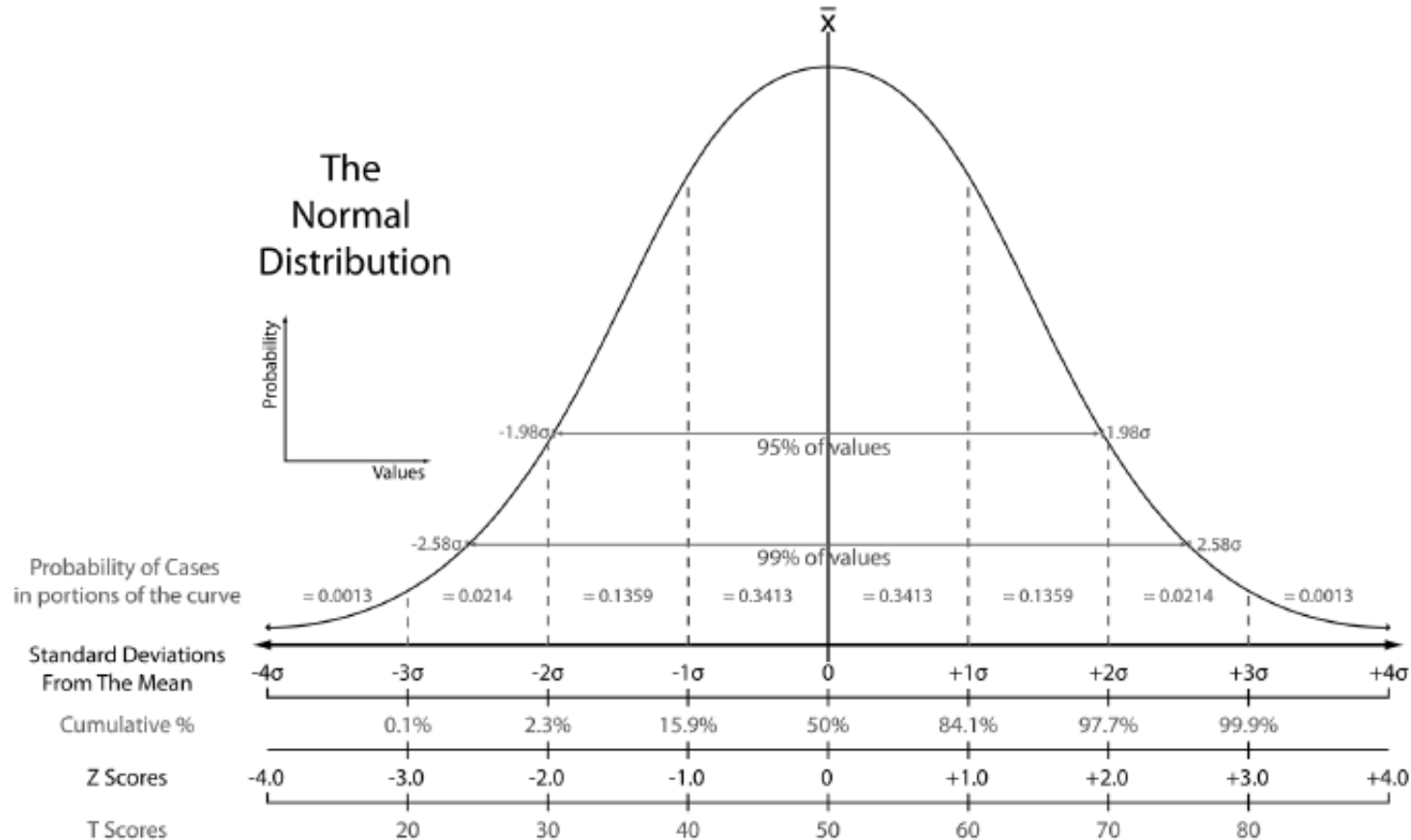
정규화된 A와 B의 유클리드 거리

$$D_{A,B} = \sqrt{(0.5 - 0.6)^2 + (0.7 - 0.5)^2 + (1.0 - 0.3)^2}$$

정규화된 자료를 이용하면, *Scale* 이 작은 변수더라도 해당 변수 내 거리의 멀고 가까움을 고려하여 군집화할 수 있게 됨

※ 참고 – Normalization

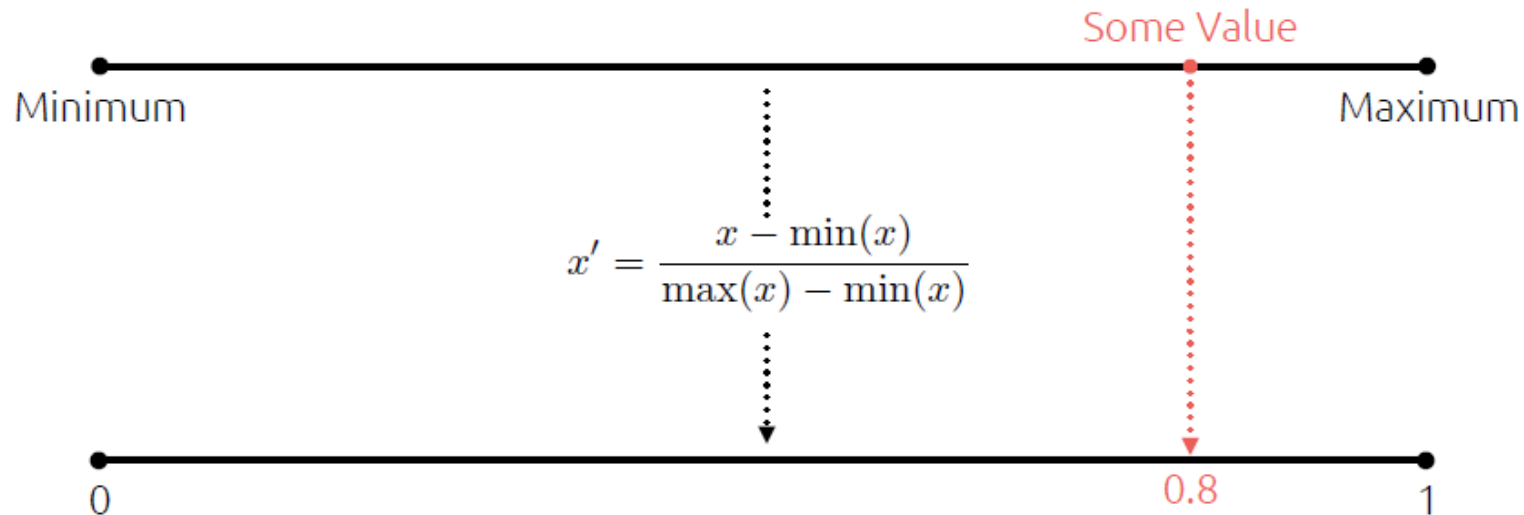
Standardization(표준화) : Z-scoring



“ $-\infty$ 에서 $+\infty$ 까지의 값을 지님 ”

※ 참고 – Normalization

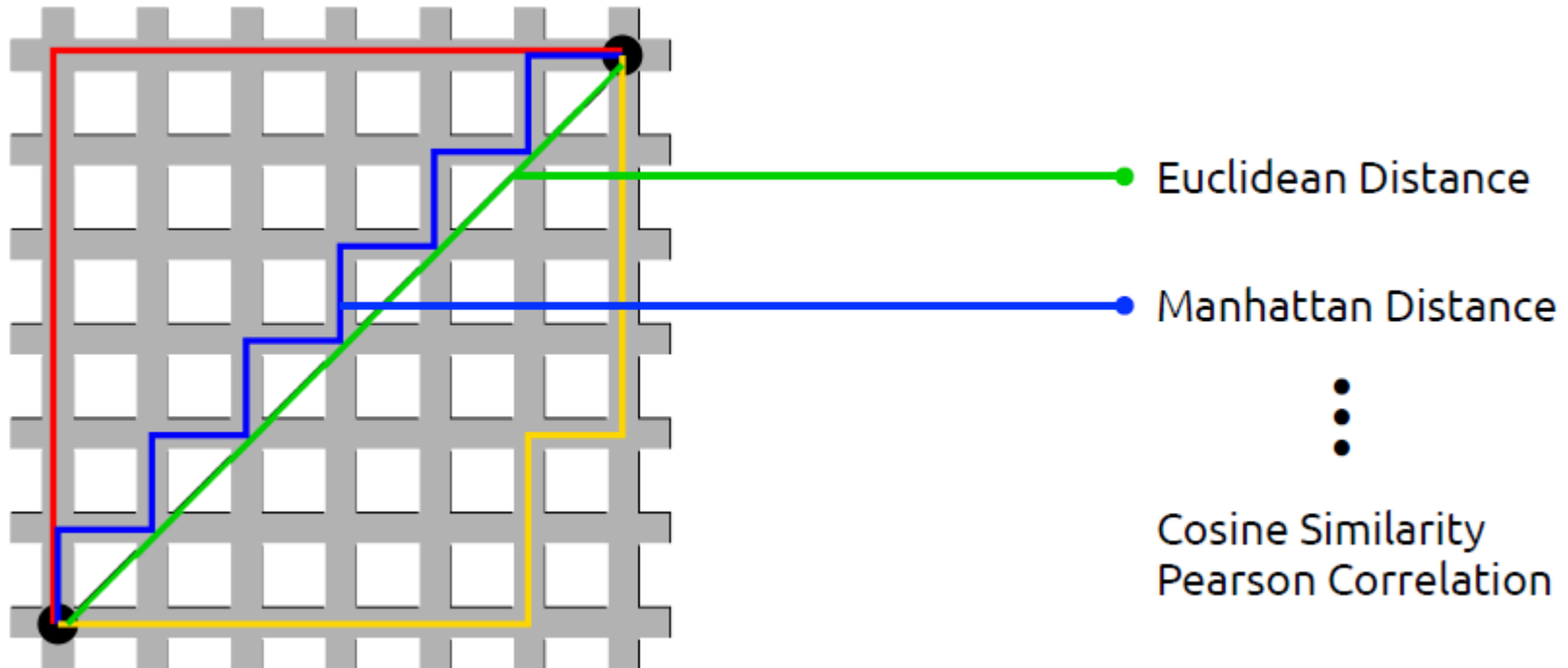
Min-Max Normalization(정규화) : Rescaling



“ 0에서 1사이의 값을 지님 ”

어떻게 근접한 이웃을 찾아갈까? – 거리(Distance)

Distance



Lecture 9-2

클러스터링 (Clustering)



Source : <https://www.freepik.com/>

*만약 새로 나온 “컴퓨터”를 팔아야 된다.
누구에게 팔아야 하겠는가?*

- 지도학습처럼 사전에 레이블(Label)이 학습된다면, 우리는 인과모형을 도출해 개개인의 특성(Feature)를 통해 자사 제품을 구매할만한 사람을 찾을 수 있음
- 하지만, 레이블(Label)이 없는 경우 의사결정 단위가 매우 커짐(왜? 개개인에 맞춰야 되기 때문!)
- 그럼 어떻게 할 수 있을까?
- 적어도 특성이 비슷한 사람끼리 묶어서 우리가 통제가능한 수준으로 의사결정 단위를 낮추면, 의사결정이 훨씬 수월할 수 있음
- 그럼 어떻게 비슷한 사람끼리 묶을 수 있을까?

클러스터링(Clustering)이란 ?

클러스터링(군집화)

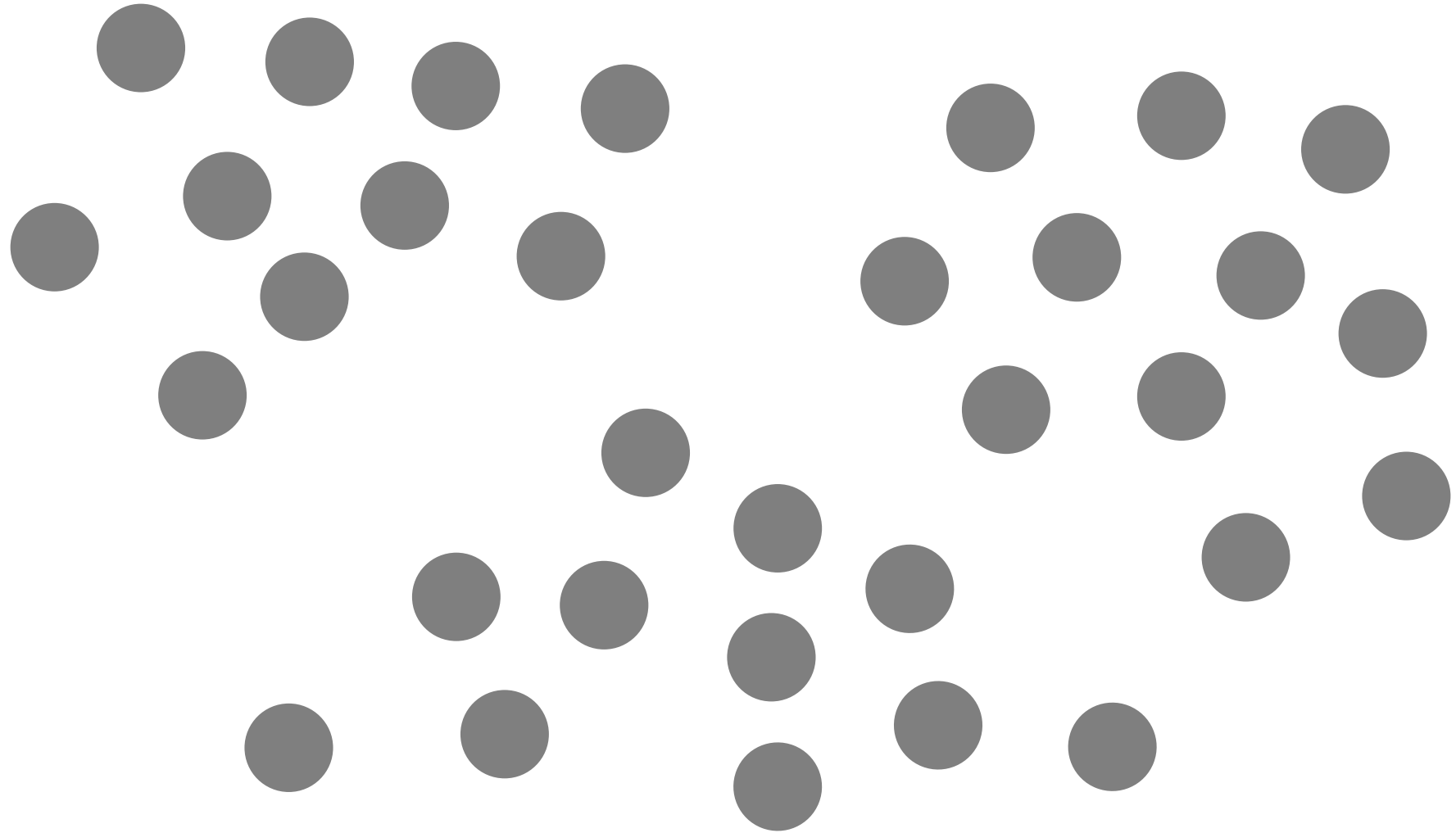


연속형인 입력변수(X)이 형성하는 특정 패턴으로부터 근접하게 분포하는 데이터들끼리 묶는 모형

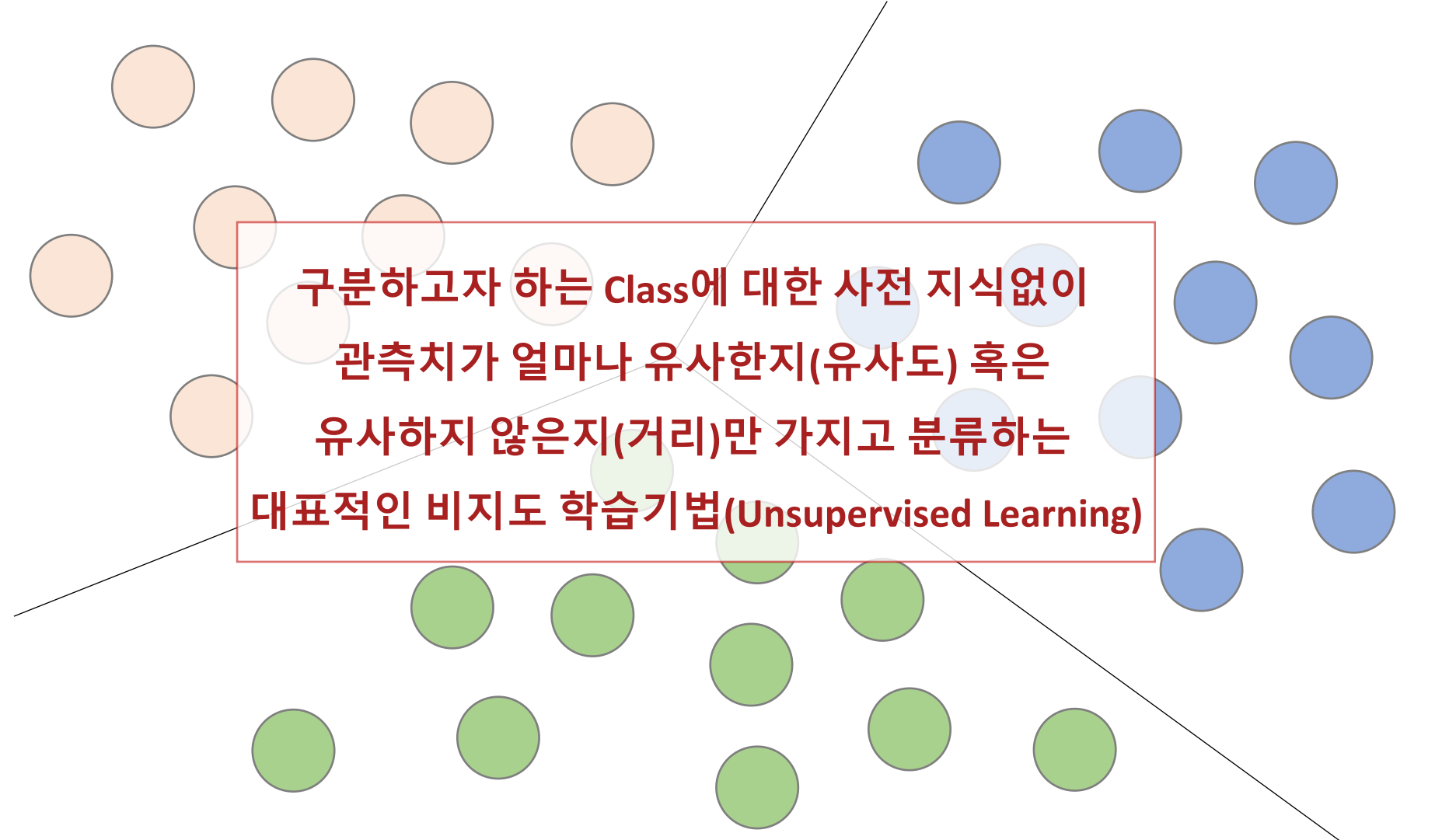
레이블이 없는 학습데이터들의 특징(Feature)만 가지고 서로 동일하거나 유사한 특징을 가진 데이터들끼리 그룹화함으로써 레이블이 없는 데이터에 레이블을 부여하는 기법

이후 새로운 데이터가 입력되면 지도학습처럼 학습한 군집을 가지고 새로운 데이터를 분류예측할 수 있음

클러스터링(Clustering)



클러스터링(Clustering)

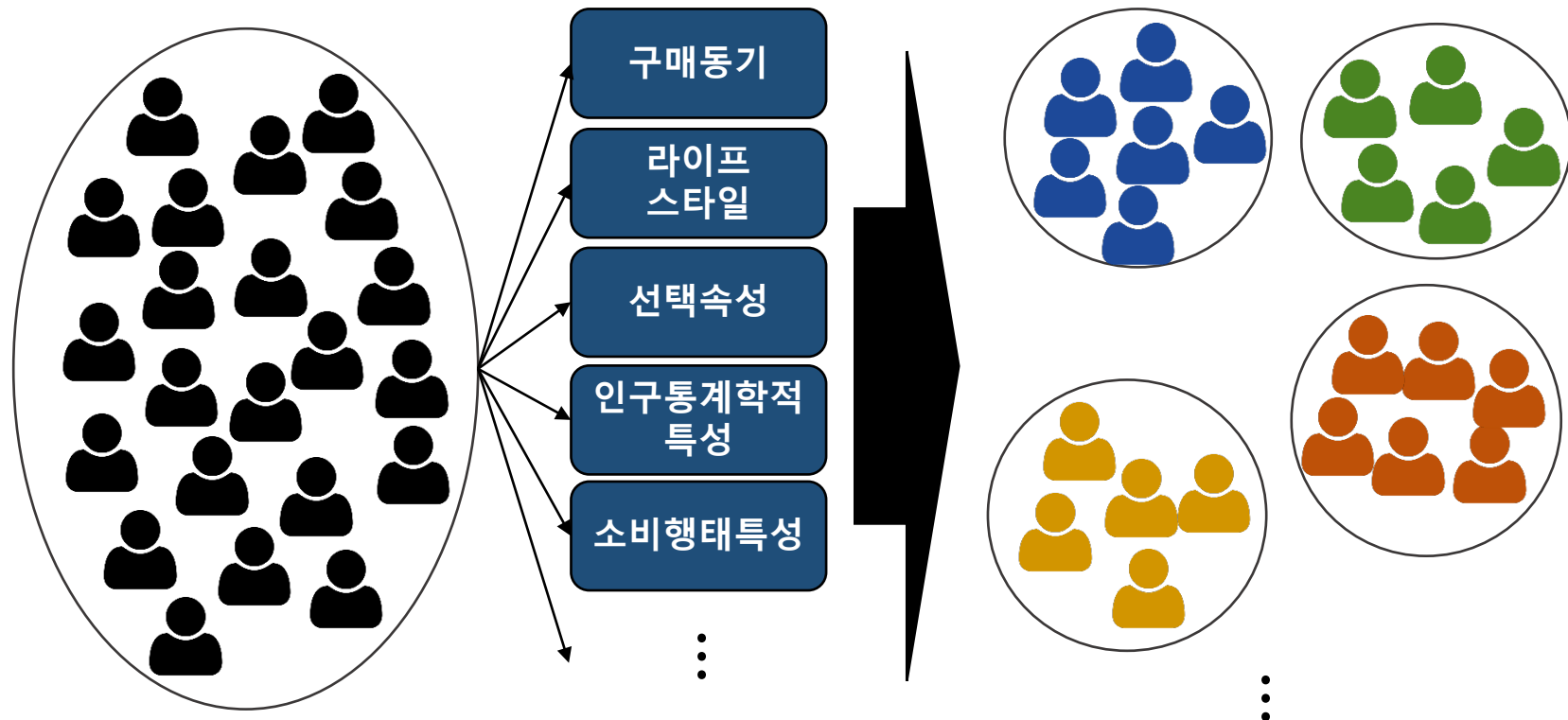


구분하고자 하는 class에 대한 사전 지식없이
관측치가 얼마나 유사한지(유사도) 혹은
유사하지 않은지(거리)만 가지고 분류하는
대표적인 비지도 학습기법(Unsupervised Learning)

시장 세분화 (Segmentation)와 클러스터링

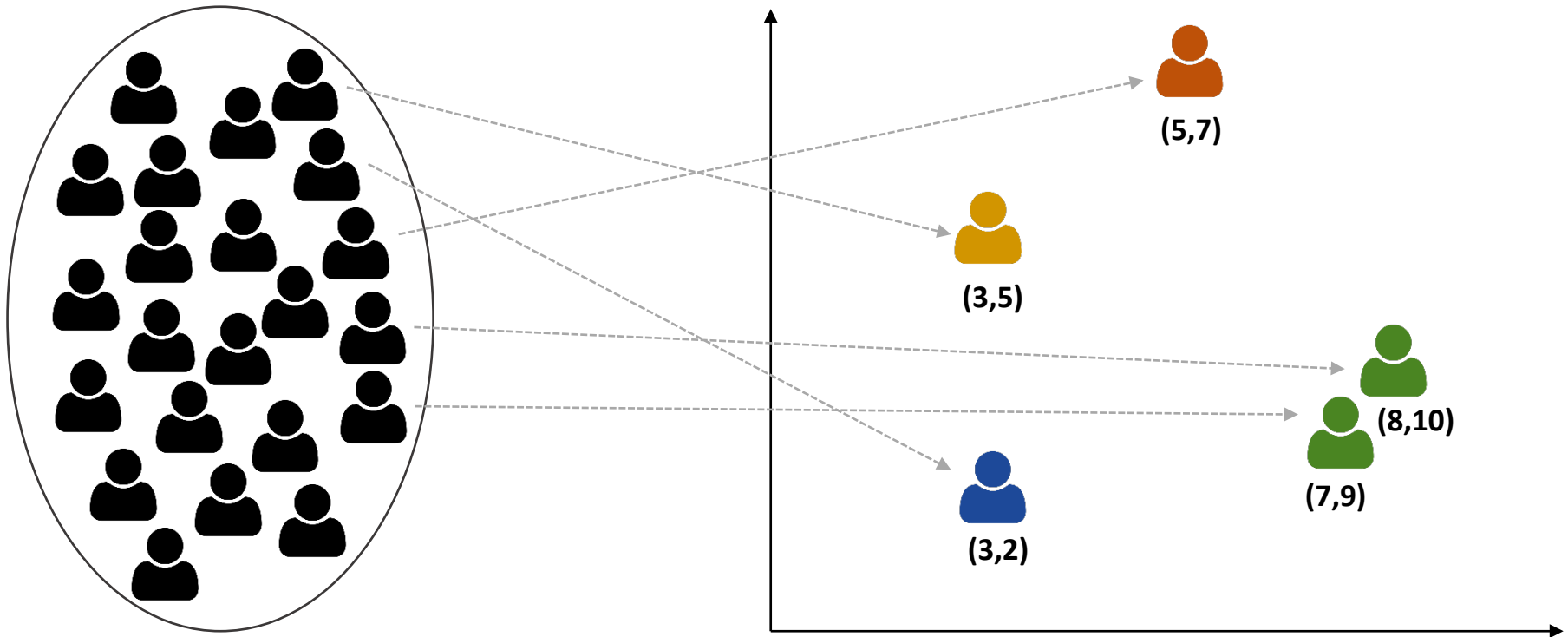
수 많은 개체를 의미있는 동질적인 집단으로 세분화하는 것은 집단 별 특성에 맞는 의사결정을 하기위해 필수적이며, 이러한 의사결정을 위해 클러스터링(Clustering)방법이 적용될 수 있음

“고객의 다양한 니즈와 특성에 따라 고객을 어떻게 구별하는 것이 타당한가 ??”



클러스터링(Clustering)의 가정

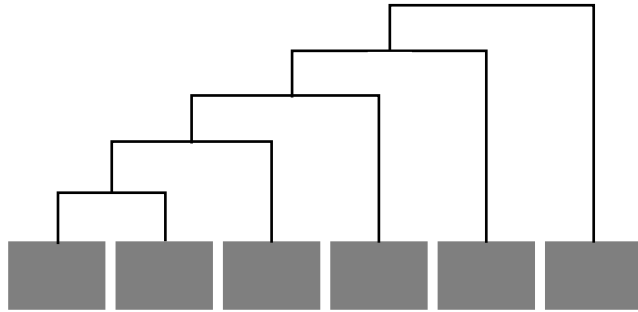
클러스터링을 하기 위해서는 각각의 개체가 좌표평면에 표현될 수 있어야 함. 즉, 군집분석을 하기 위해서 투입되는 변수들이 각각 연속형의 실수값을 지녀야 함



“ 각 개체의 변수가 갖는 실수값을 바탕으로 좌표평면에 표시함 ”

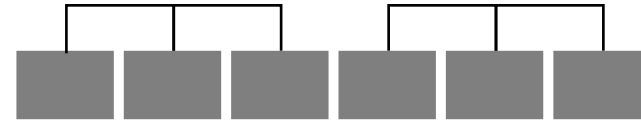
대표적인 Clustering algorithm

계층적 군집분석 (Hierarchical clustering)



- 계층적 군집분석은 각각의 개체들을 순차적 및 계층적으로 유사한 개체끼리 묶어 나가는 알고리즘
- 군집 수를 사전에 정의할 필요 없음
- Tree 구조를 통해 계층 구조를 살펴볼 수 있음

K-means 군집분석 (K-means clustering)

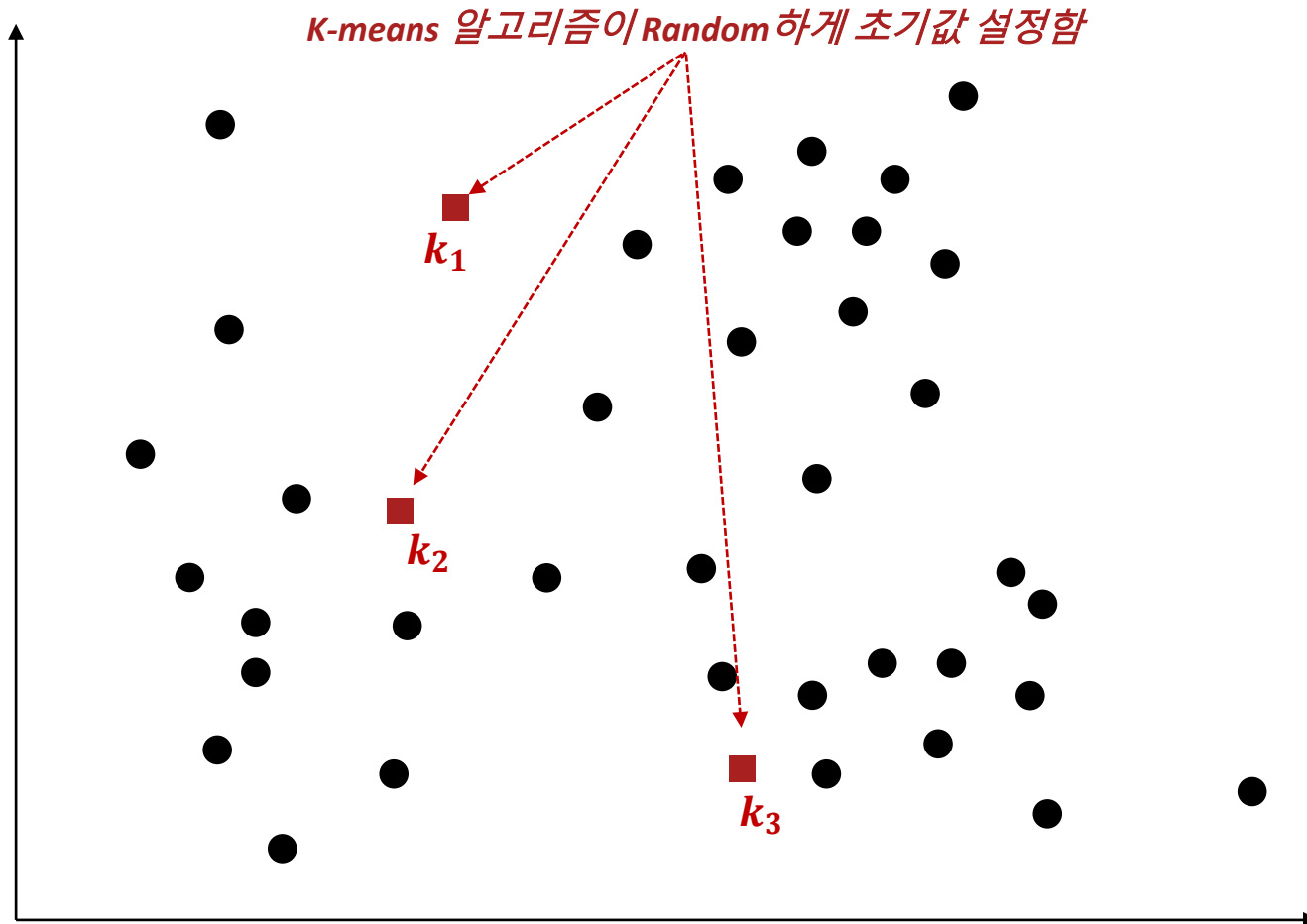


- k 개의 집단으로 분류하되, 집단 간 거리는 최대한 멀고, 집단 내 개체 간 거리는 최대한 가깝도록 군집화(clustering)하는 알고리즘
- 계층적 군집분석과 달리 k 의 값을 직접 설정해줘야 함

K-means 클러스터분석 (Cluster Analysis)

예 : $k = 3$ 으로 설정

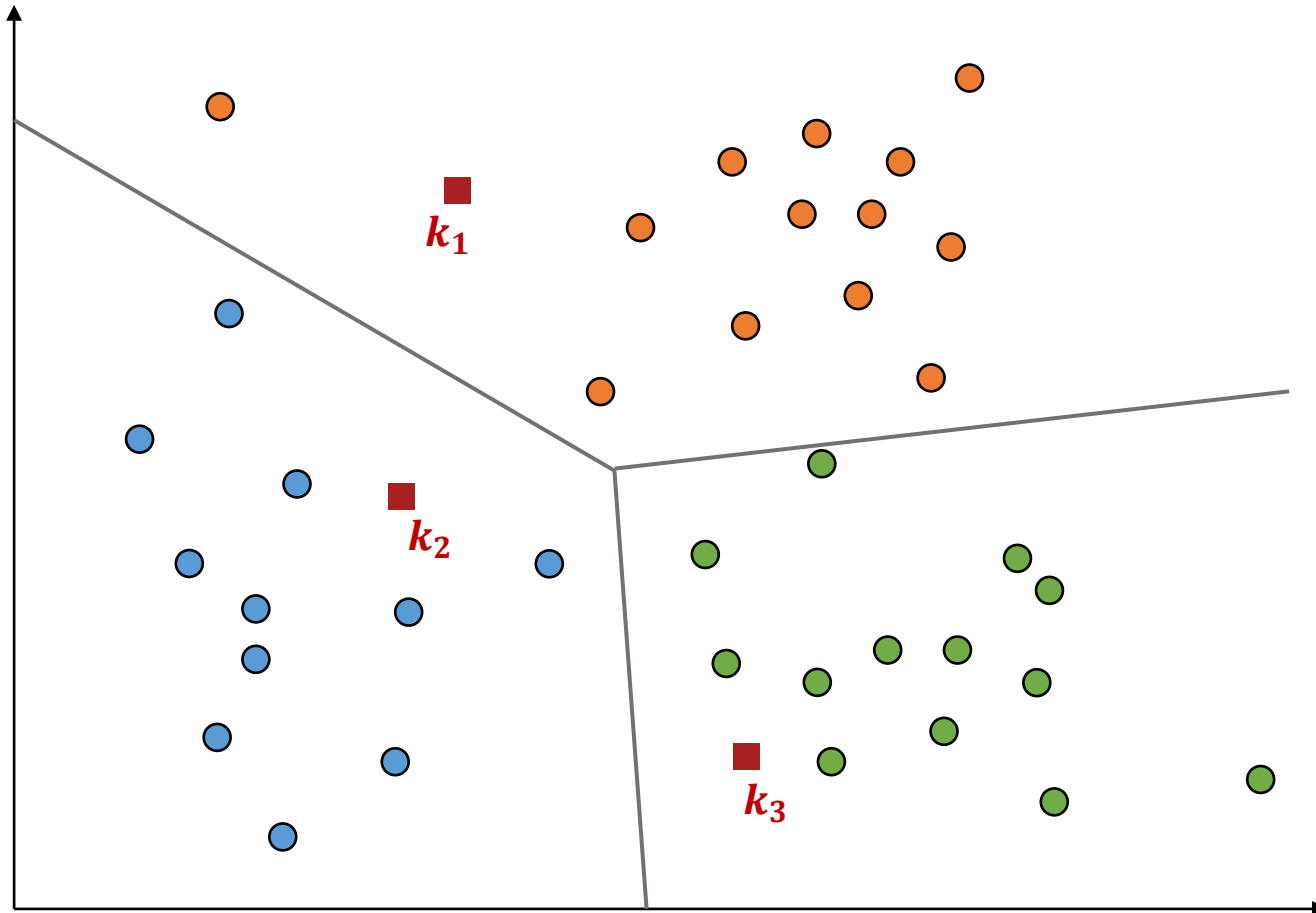
1) 초기 k 개의 중심(Centroid)이 random하게 선택됨



K-means 클러스터분석 (Cluster Analysis)

예 : $k = 3$ 으로 설정

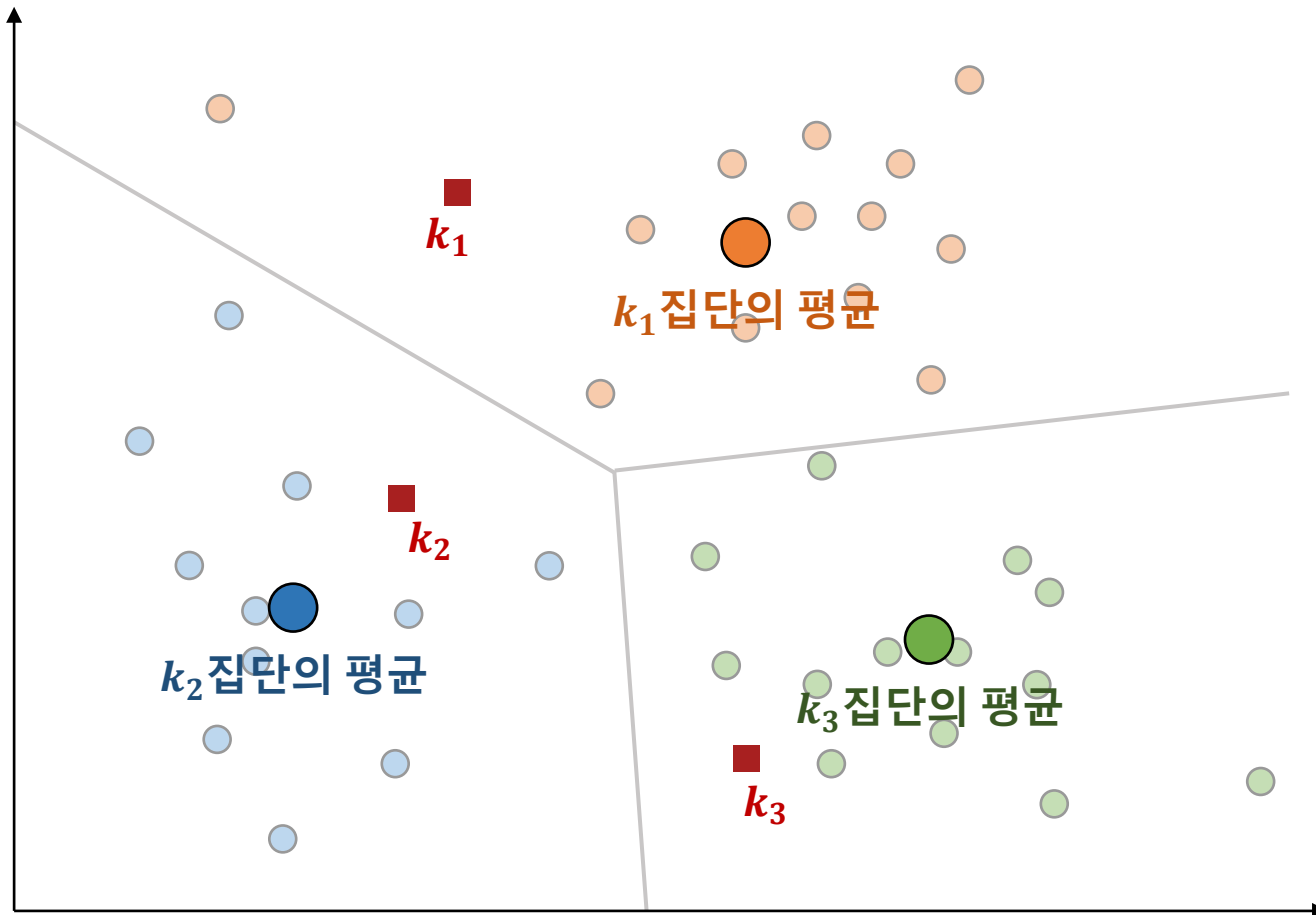
2) 각 개체(Observation)는 자신에게 가장 가까운 초기 중심(centroid) k 에 할당됨



K-means 클러스터분석 (Cluster Analysis)

예 : $k = 3$ 으로 설정

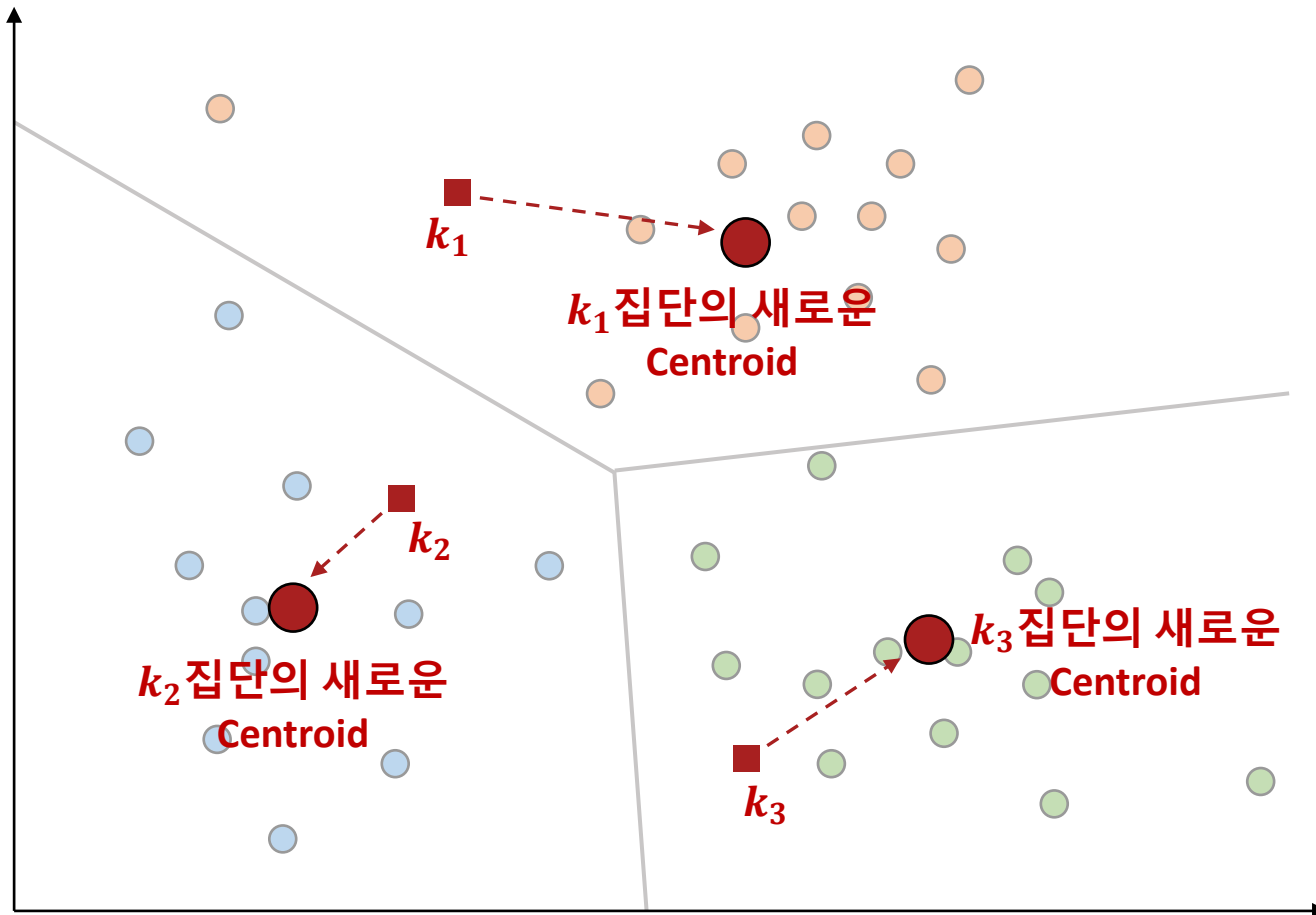
3) 같은 Centroid에 할당된 개체들의 평균(Mean)을 구한다



K-means 클러스터분석 (Cluster Analysis)

예 : $k = 3$ 으로 설정

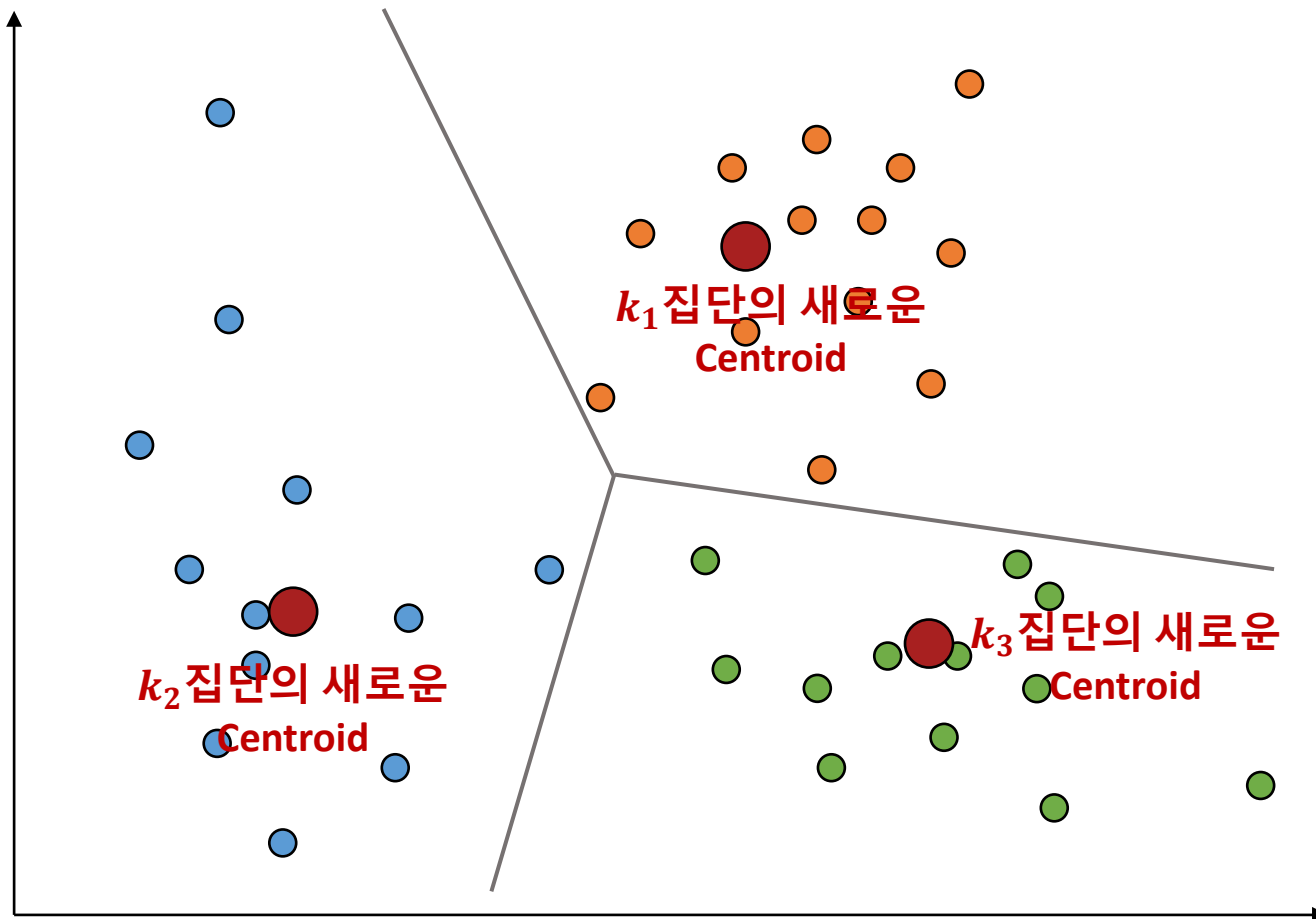
4) 초기 설정된 Centroid는 없어지고, 새롭게 계산한 각 집단의 평균으로 Centroid가 이동함



K-means 클러스터분석 (Cluster Analysis)

예 : $k = 3$ 으로 설정

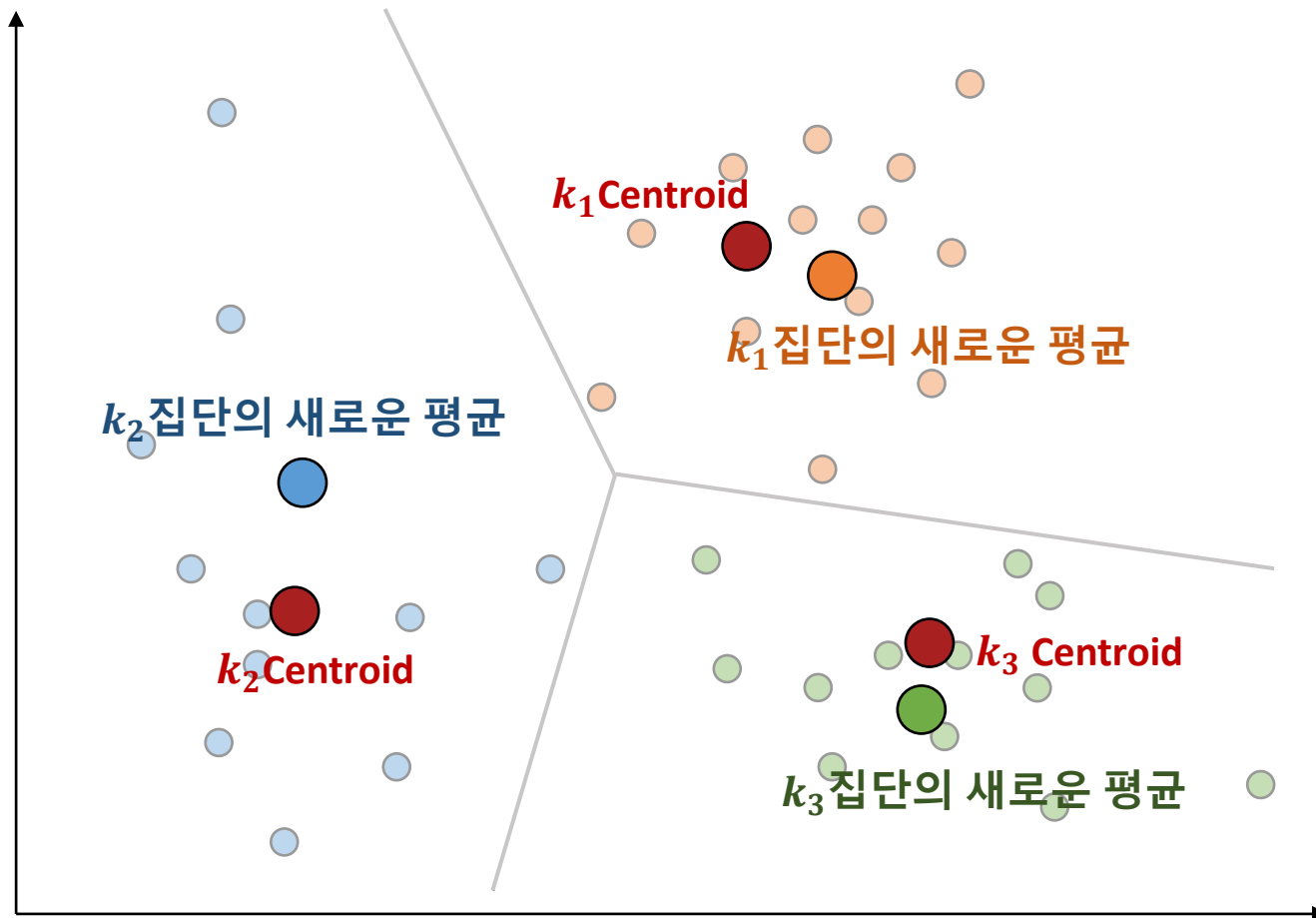
5) 새롭게 정의된 Centroid를 기준으로 각 개체는 가장 가까이에 있는 Centroid에 다시 속하도록 함



K-means 클러스터분석 (Cluster Analysis)

예 : $k = 3$ 으로 설정

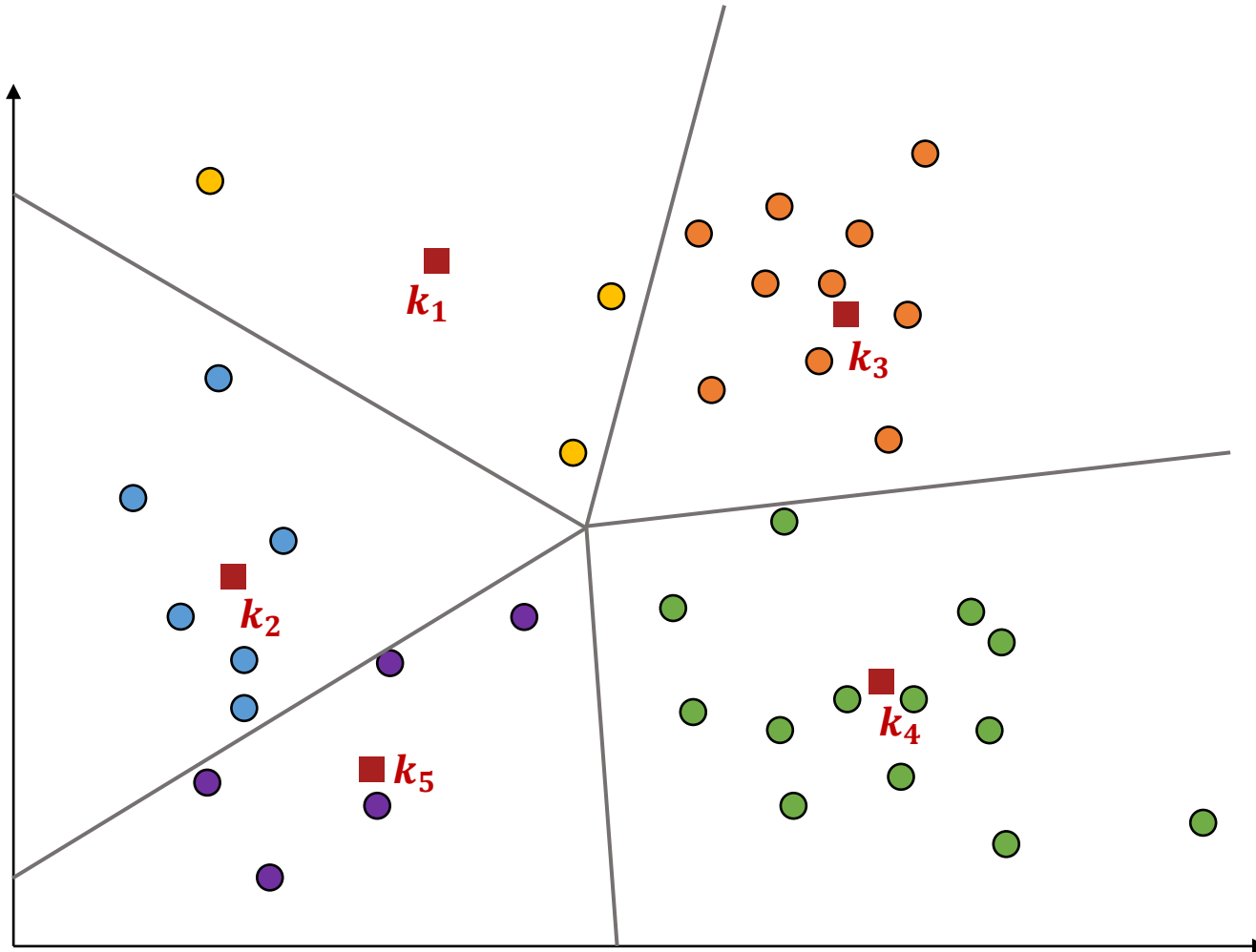
6) 1)부터 5)까지의 과정을 반복적으로 시행하되, 더 이상 집단이 바뀌지 않을 때, 반복을 멈춤



K-means 클러스터분석 (Cluster Analysis)

예 : $k = 5$ 으로 설정

마찬가지로, 초기값을 랜덤하게 설정 후 반복을 통해 군집이 바뀌지 않을 때 멈춤



K-Nearest Neighbor

K-Means 클러스터링

잠재계층 분석

- ✓ 관측치를 그룹핑해야 되는데, 몇 개의 그룹이 존재하는지 Label이 없어서 모른다.
- ✓ 따라서, “**그룹의 수 혹은 계층의 수**” 자체가 **하이퍼 파라미터 (Hyperparameter)**가 되며, 우리는 그룹의 수를 변화시켜 가면서 모형을 돌려서 최적의 모형을 찾는 “파라미터 탐색” 과정을 거쳐야 한다.

※ 하이퍼 파라미터 (Hyperparameter)

- 모형 내에서 결정되는 파라미터가 아닌, 학습 알고리즘 자체의 성능을 결정하는 파라미터로, 학습전에 미리 사용자에게 의해 정의되는 파라미터를 의미함.
- 대표적으로 군집분석 및 KNN의 “K”, 잠재계층분석의 “class 수” 등이 있다.

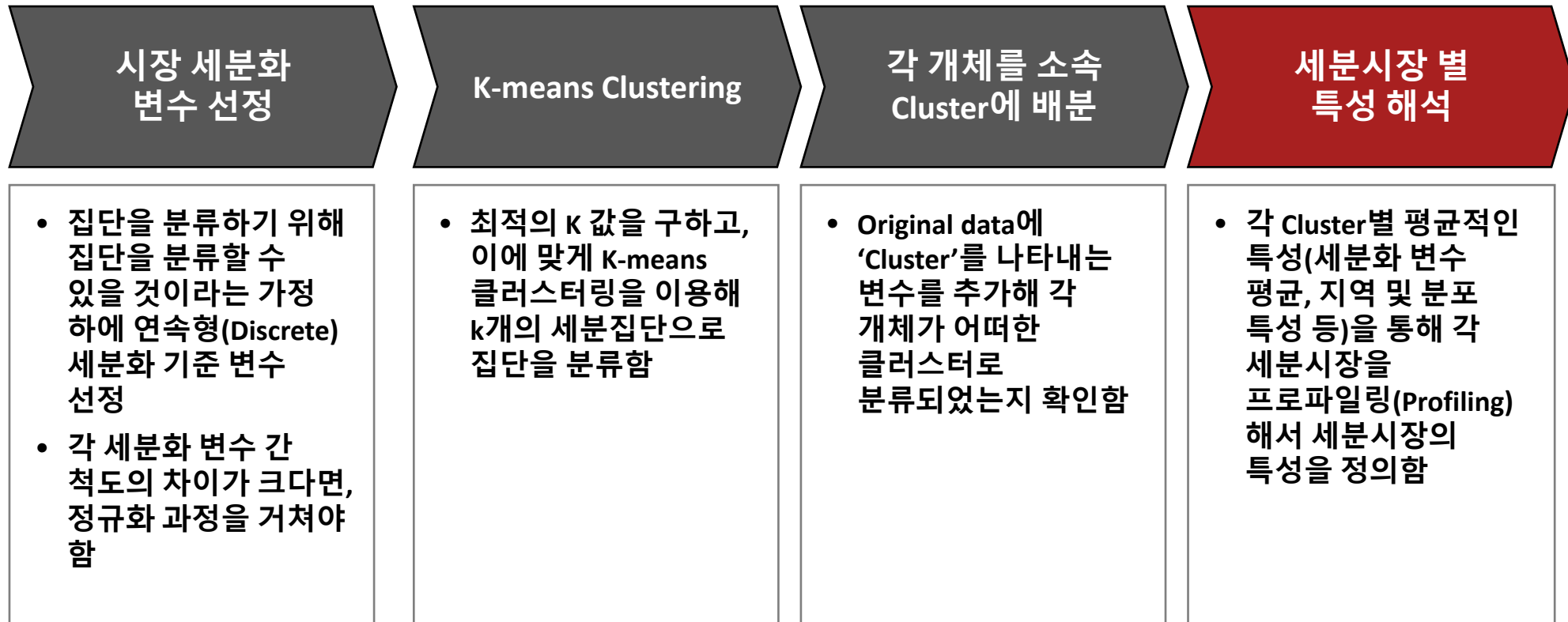
K-means Algorithm의 한계

K-means 클러스터링의 한계

- ✓ 모든 데이터에 대해 사용할 수 없음. 데이터의 경계가 선형으로 분리되지 않는 자료에 대해서는 사용 불가
- ✓ 군집화 변수가 반드시 실수값을 갖는 연속형 변수여야 함
- ✓ 이상치(outlier)가 있는 경우, 초기값에 매우 민감하게 반응하고, 분석할 때마다 결과가 조금씩 달라짐

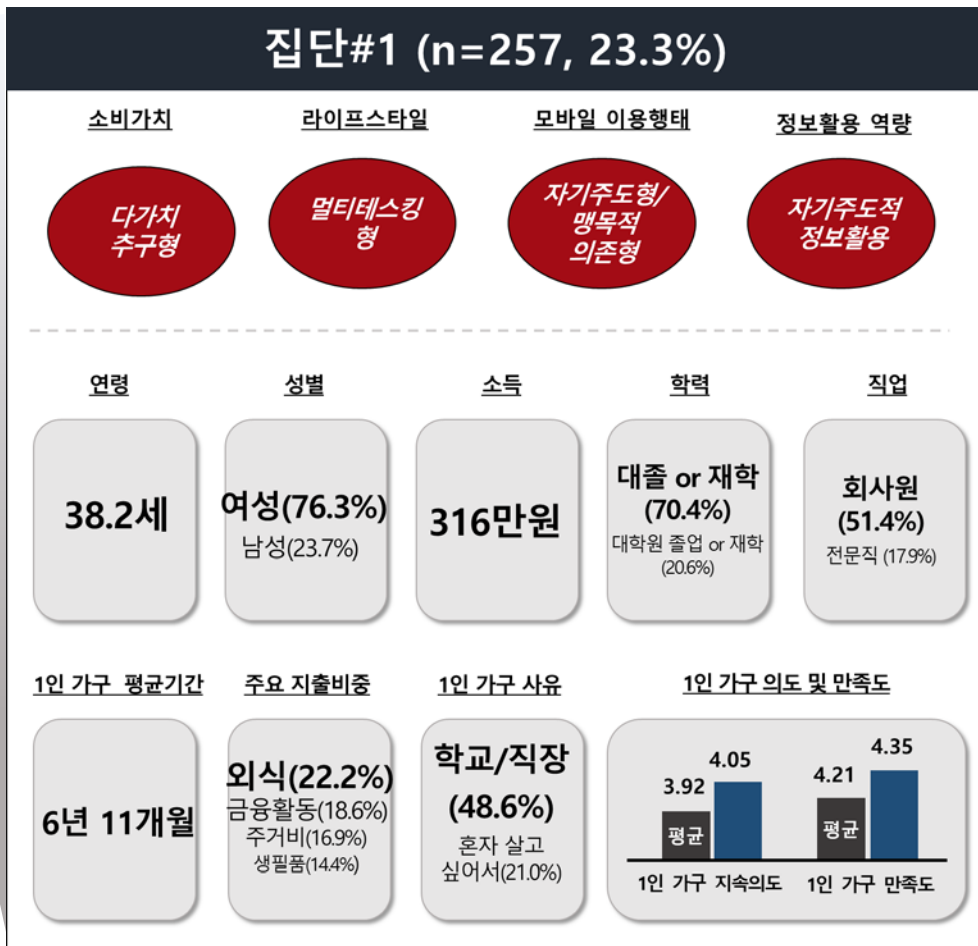
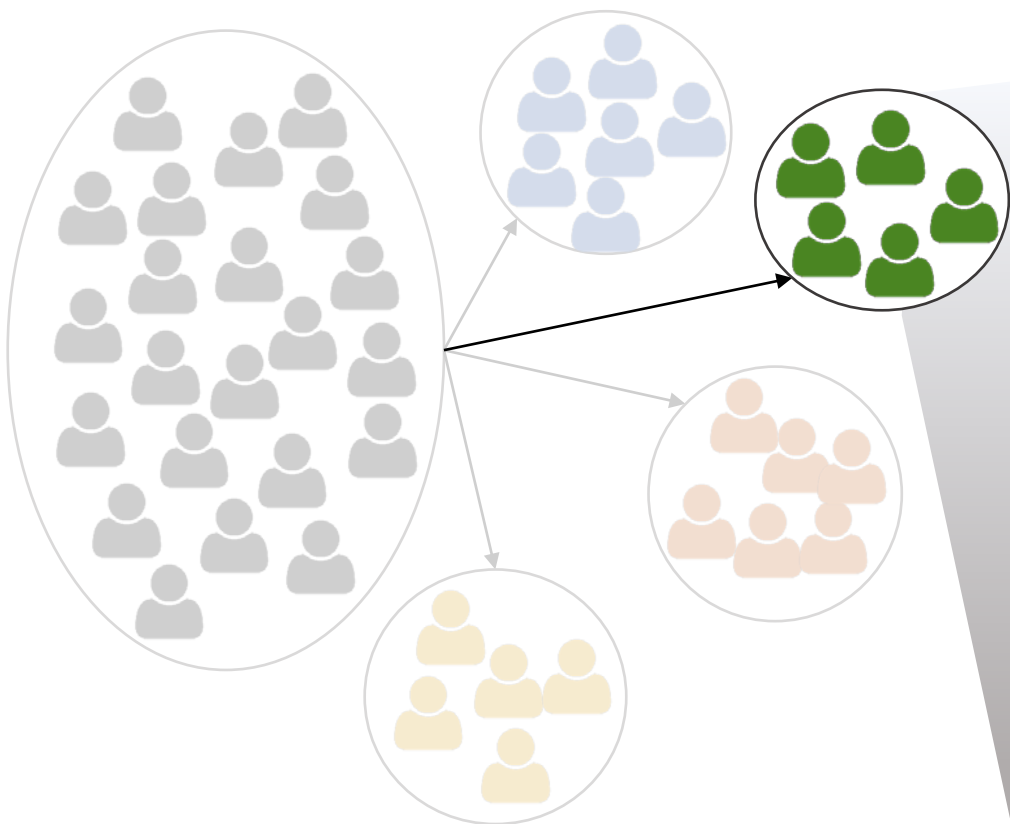
클러스터링을 이용한 비지도 분류 결과 적용 방안

클러스터링을 실무적으로 활용하기 위해서는 방법론적 접근도 중요하지만, 더욱 중요한 것은 세분화된 각 세분집단을 Profiling하는 것임



세분시장 별 특성 해석 ; 프로파일링(Profiling)

군집별 프로파일링은 Subset의 요약통계량을 구하는 과정이라 볼 수 있음. 즉, 특정 집단을 Filtering한 후 각 집단의 특성을 파악하는 과정을 프로파일링이라 함



Lecture 9-3

잠재계층분석
(Latent Class
Model)

Classification by discrete variable



집단 분류를 반드시 실수값을 갖는 연속형 변수로
해야되는가?

성별, 직업, 행동유형 등 이산형(Discrete)한 변수로 집단을
분류할 수는 없는가?

잠재계층분석(Latent Class Analysis; LCA)

잠재계층분석은 이항형(Binary) 변수 혹은 다항형(polytomous) 항목에 대한 반응을 기반으로 유사한 반응집단을 분류하는 기법으로 이산형 변수에 대한 집단세분화를 위해 적용할 수 있음

잠재계층분석(LCA) 분석 Flow

집단세분화 변수 선정

- 집단을 분류하기 위해 이산형(Discrete)의 세분화 기준 변수 선정
- 이산형 변수의 항목 수는 **5개를 넘지 않도록 하며**, 넘는 경우 일부 항목의 통합 등을 통해 조절

최적 세분집단(Class) 수의 결정

- 최적 세분시장 수를 결정하기 위해 AIC와 BIC 값을 이용하며, AIC와 BIC 값이 가장 작은 모형이 최적의 모형
- 이때, 최적의 모형의 집단(class) 수가 최적 세분집단 수

항목반응행렬 (Item-response matrix) 도출

- 항목반응행렬은 각 집단의 각 변수에 대한 반응확률을 나타낸 행렬임
- 항목반응행렬을 통해 각 집단의 특성을 정의할 수 있음

세분시장 별 특성 해석

- 항목반응행렬을 통해 정의된 집단별 특성을 해석함
- 각 집단별로 평균적인 특성을 추가적으로 분석함으로써 집단에 대한 Profiling 실시

예시 : 외식소비자의 세분화

Step 1. 시장세분화 기준 변수 선정

변 수	측 정	범주 수	변 수	측 정	범주 수
외식 소비액	1. 0초과~30,000원 이하	4개	외식 장소	1. 한식당(육류 제외)	8개
	2. 30,000초과 ~100,000원 이하			2. 육류전문점	
	3. 100,000초과~300,000원 이하			3. 일식(횃집 포함)	
	4. 300,000초과			4. 중식	
외식 빈도	1. 매일	6개		5. 서양식	
	2. 주 4~5회			6. 치킨전문점	
	3. 주 2~3회			7. 패스트푸드/분식	
	4. 주 1회			8. 기타(기타 외국식, 뷔페레스토랑)	
	5. 2주에 1회		외식 선택 기준	1. 음식 외적요소 (청결도, 가격, 서비스, 분위기, 메뉴다양성)	3개
	6. 1달에 1회이하				
외식 동기	1. 맛있는 음식을 즐기기 위해	4개	3. 편의성 (교통편의, 예약용이, 부대시설)		
	2. 근로, 학업 등의 사유				
	3. 식사준비가 귀찮아서				
	4. 특별한 날				

Source : 최규완 · 방도형(2017). 외식소비시장 시장세분화 연구, 관광학연구

예시 : 외식소비자의 세분화

Step 2. 최적 시장 수의 결정

클래스 수	AIC	BIC
1	66549.20	66681.00
2	63894.27	64164.47
3	62992.84	63401.42
4	62730.45	63277.43
5	62551.99	63237.36
6	62463.61	63287.37
7	62407.87	63370.02
8	62392.06	63492.60
9	62342.90	63581.83
10	62318.24	63695.56

- AIC(Akaike-information criterion)과 BIC(Bayesian-information criterion)을 기준으로 두 정보량이 **최저값이 되는 점이 최적의 모형**이 됨
- 일반적으로 LCA와 관련된 연구들에서 AIC 보다 BIC가 더 적합한 정보기준량이라 밝히고 있음
- 실제로 분석결과를 살펴보면 AIC는 최적값으로 수렴하지 않고 지속적으로 줄어드는 것을 알 수 있음
- 따라서, LCA를 시행할 시 AIC와 BIC가 다른 결과를 나타낸다면, BIC 값을 우선적으로 고려하는 것이 적절함
- 본 예시에서는 최적 세분시장의 수는 “5개”로 나타남

예시 : 외식소비자의 세분화

Step 3. 항목반응행렬 도출

구분		집단 1	집단 2	집단 3	집단 4	집단 5
세분시장 별 크기		0.164	0.103	0.136	0.318	0.107
외식소비액	0원 ~ 30,000원 이하	0.123	0.203	0.029	0.000	0.821
	30,000원 초과 ~ 100,000원 이하	0.755	0.648	0.132	0.238	0.172
	100,000원 초과 ~ 300,000원 이하	0.122	0.143	0.816	0.752	0.004
	300,000원 초과	0.000	0.006	0.022	0.011	0.003
외식빈도	매일	0.001	0.006	0.130	0.000	0.000
	주 4~5회	0.010	0.088	0.731	0.053	0.002
	주 2~3회	0.023	0.380	0.092	0.527	0.002
	주 1회	0.350	0.331	0.004	0.413	0.000
	2주에 1회	0.397	0.127	0.023	0.007	0.169
	1달에 1회 이하	0.218	0.068	0.020	0.000	0.827
외식동기	맛있는 음식을 즐기기 위해	0.537	0.422	0.164	0.507	0.313
	근로, 학업 등	0.054	0.284	0.757	0.165	0.046
	식사준비 귀찮아서	0.077	0.220	0.063	0.053	0.086
	특별한 날	0.332	0.074	0.016	0.274	0.555
외식장소	한식당(육류 제외)	0.340	0.456	0.613	0.300	0.530
	육류전문점	0.459	0.167	0.240	0.551	0.329
	일식(횃집 포함)	0.099	0.009	0.011	0.069	0.028
	중식	0.013	0.015	0.020	0.003	0.028
	서양식	0.058	0.004	0.003	0.028	0.017
	치킨전문점	0.010	0.070	0.006	0.036	0.017
	패스트푸드/분식	0.004	0.267	0.105	0.000	0.038
	기타(기타 외국식, 뷔페)	0.017	0.013	0.002	0.013	0.013
외식 선택기준	음식외적요소(청결도, 가격, 서비스, 분위기, 메뉴다양성)	0.407	0.527	0.359	0.402	0.466
	음식요소(건강한요리, 음식의맛, 음식의양)	0.560	0.449	0.630	0.522	0.498
	편의성(교통편의, 예약용이, 부대시설)	0.034	0.024	0.010	0.077	0.037

Source : 최규완 · 방도형(2017). 외식소비시장 시장세분화 연구, 관광학연구

Lecture 9-4

Clustering 과
LCA를 이용한
2단계 세분화

1인 가구 소비행태 조사 설계

1인 가구 소비 특성 및 소비 행태를 파악하기 위해 다음과 같이 조사 설계함

"1인 가구 증가에 따른 소상공인 대응 방안 연구"

조사 대상	대한민국 국적의 현재 혼자 거주 중인 가구원	
조사 기간	2020. 07.20 ~07.24	
유효 표본 수	1,104부	
주요 조사 항목	<ul style="list-style-type: none">• 인구통계학적 특성• 1인 가구 소비 특성• 1인 가구 소비가치	<ul style="list-style-type: none">• 1인 가구 라이프스타일• 1인 가구 모바일 이용행태• 1인 가구 정보활용 역량
조사 대행기관	(주)마크로밀엠브레인	

조사 대상 인구통계학적 특성

구분	항목	N	비율(%)	구분	항목	N	비율(%)
성별	남자	351	31.8	월 평균 소득	소득 없음	80	7.2
	여자	753	68.2		1원 이상 - 50만원 미만	24	2.2
	총계	1,104	100.0		50만원 이상 - 100만원 미만	41	3.7
연령 (평균: 34.92세)	만18-19세	5	0.5		100만원 이상 - 150만원 미만	37	3.4
	만20-29세	372	33.7		150만원 이상 - 200만원 미만	202	18.3
	만30-39세	443	40.1		200만원 이상 - 250만원 미만	232	21.0
	만40-49세	172	15.6		250만원 이상 - 300만원 미만	207	18.8
	만50-59세	95	8.6		300만원 이상 - 350만원 미만	121	11.0
	만60-69세	17	1.5		350만원 이상 - 400만원 미만	58	5.3
	전체	1,104	100.0		400만원 이상 - 450만원 미만	30	2.7
혼인	미혼	1,055	95.6		450만원 이상 - 500만원 미만	33	3.0
	기혼	49	4.4		500만원 이상	39	3.5
	전체	1,104	100.0		전체	1,104	100.0
최종 학력	중학교 졸업 이하	7	0.6	거주 지역	서울	482	43.7
	고등학교 졸업	146	13.2		부산	64	5.8
	대학교 졸업(재학)	760	68.8		대구	47	4.3
	대학원 졸업(재학)	191	17.3		인천	49	4.4
	전체	1,104	100.0		광주	25	2.3
직업	대학(원)생	86	7.8		대전	39	3.5
	회사원	548	49.6		울산	6	0.5
	전문직	164	14.9		경기도	200	18.1
	서비스직	93	8.4		강원도	23	2.1
	자영업	57	5.2		충청북도	27	2.4
	공무원	43	3.9		충청남도	24	2.2
	무직	59	5.3		전라북도	21	1.9
	기타	54	4.9		전라남도	9	0.8
	전체	1,104	100.0		경상북도	35	3.2
					경상남도	37	3.4
					제주도	10	0.9
					세종	6	0.5
					전체	1,104	100.0

1인 가구 주요 특성

구분	항목	N	비율(%)	구분	항목	N	비율(%)
혼자 사는 이유	학교/직장 때문에	468	42.4	추후 혼자 살 기간	1년 미만	47	4.3
	혼자 살고 싶어서	237	21.5		1년 - 2년 미만	144	13.0
	배우자와 이혼, 별거, 사별로	69	6.3		2년 - 5년 미만	299	27.1
	미혼이라서	282	25.5		5년 - 10년 미만	105	9.5
	가족의 학업, 직장, 이민 등의 이유로	42	3.8		10년 이상	242	21.9
	기타	6	0.5		잘 모르겠다/불확실하다	267	24.2
	전체	1,104	100.0		전체	1,104	100.0
추후 혼자 살 의향	전혀 그렇지않다	36	3.3	주거 형식	자가	153	13.9
	그렇지 않다	122	11.1		전세	402	36.4
	그저 그렇다	118	10.7		월세	448	40.6
	그런 편이다	393	35.6		전월세 혼합	101	9.1
	매우 그렇다	385	34.9		전체	1,104	100.0
	어쩔 수 없이 혼자 살 수밖에 없다	50	4.5	주거 형태	아파트	255	23.1
	전체	1,104	100.0		단독주택	141	12.8
만족도	전혀 그렇지않다	13	1.2		오피스텔	342	31.0
	그렇지않다	39	3.5		기타	366	33.2
	그저 그렇다	138	12.5		전체	1,104	100.0
	그런 편이다	445	40.3	주거 지역	주거지 (아파트, 연립, 단독 밀집지역)	932	84.4
	매우 그렇다	460	41.7		상업지	130	11.8
	어쩔 수 없이 혼자 살 수밖에 없다	9	0.8		농, 어촌 지역	25	2.3
	전체	1,104	100.0		기타	17	1.5
					전체	1,104	100.0

군집분석을 위한 조사 문항

군집분석을 위해 소비가치, 라이프스타일, 모바일 이용행태, 정보활용역량 각각의 문항을 이용해 군집분석을 실시함

내용		5점 척도				
		매우 그렇지 않다	그렇지 않다	보통 이다	그렇다	매우 그렇다
소비 가치	8-1	나는 소비를 결정할 때, 사람들이 다른 사람들의 반응을 중요하게 고려하는 편이다.				
	8-2	나는 남들보다 새로운 것을 먼저 소비하고자 하는 편이다.				
	8-3	나는 소비할 때 환경적 요소를 중요하게 고려하는 편이다.				
	8-4	나는 소비할 때 당장 필요하지 않더라도 분위기 및 감정에 따라 구매하는 편이다.				
	8-5	나는 소비할 때, 가격이 저렴한 물건을 자주 사는 것보다 비싸도 한번 사서 오래 쓰는 것을 선호한다.				
	8-6	나는 작은 물건을 살 때에도 브랜드 및 상표를 중요시 하는 편이다.				
	8-7	나는 작은 물건을 살 때에도 가격을 철저히 비교한 후 구매한다.				

내용		5점 척도				
		매우 그렇지 않다	그렇지 않다	보통 이다	그렇다	매우 그렇다
라이프 스타일	9-1	나는 외모나 미용에 관심이 많은 편이다.				
	9-2	나는 유행에 민감한 편이다.				
	9-3	나는 안정적인 생활보다 모험적인 생활을 추구하는 편이다.				
	9-4	나는 친구 또는 동창회 등의 모임에 자주 나가는 편이다.				
	9-5	나는 문화 및 취미활동에 적극적으로 참여하고 있다.				
	9-6	나는 집에서 하는 활동을 선호하는 편이다.				
	9-7	나는 사회가 변하는 속도가 빨라 종종 부담감을 느낄 때가 있다.				
	9-8	나는 새로운 제품이나 기술에 대해 적극적으로 수용하는 편이다.				

내용		5점 척도				
		매우 그렇지 않다	그렇지 않다	보통 이다	그렇다	매우 그렇다
모바일 이용 행태	10-1	나는 모바일/온라인 쇼핑을 오프라인 쇼핑보다 선호한다.				
	10-2	나는 귀찮더라도 오프라인으로 확인하고, 모바일/온라인으로 구매하는 편이다.				
	10-3	나는 작은 물건을 사더라도 오프라인 보다 모바일/온라인으로 구매하는 편이다.				
	10-4	나는 모바일/온라인 쇼핑이 가격을 비교하기 편해서 이용한다.				
	10-5	나는 모바일/온라인 쇼핑이 다른 사람들의 의견을 참고하기 편해서 이용한다.				
	10-6	나는 모바일/온라인 쇼핑이 오프라인으로 대면하는 것보다 덜 부담스러워서 이용한다.				
	10-7	나는 모바일/온라인 쇼핑이 구매할 수 있는 대안이 더 많아서 선호한다.				

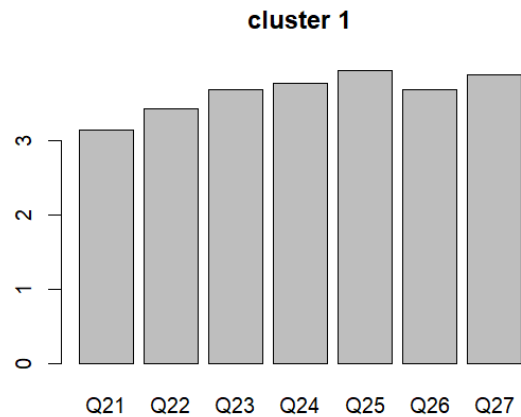
내용		5점 척도				
		매우 그렇지 않다	그렇지 않다	보통 이다	그렇다	매우 그렇다
정보 활용 역량	11-1	나는 무언가 구매할 때, 필요한 정보를 검색하는 방법을 잘 알고 있다.				
	11-2	나는 무언가 구매할 때, 다른 사람의 의견을 신중하게 참고하는 편이다.				
	11-3	나는 합리적인 소비를 위해 정확한 정보를 활용하여 바람직한 소비를 하고 있다고 생각한다.				
	11-4	나는 소비 시 직접 정보를 찾아보기 보다는 판매자의 추천에 의존하는 편이다.				
	11-5	나는 정보가 너무 많아 때때로 구매 결정이 오히려 어렵다고 느낄 때가 있다.				

1인 가구 특성에 따른 군집분석 결과

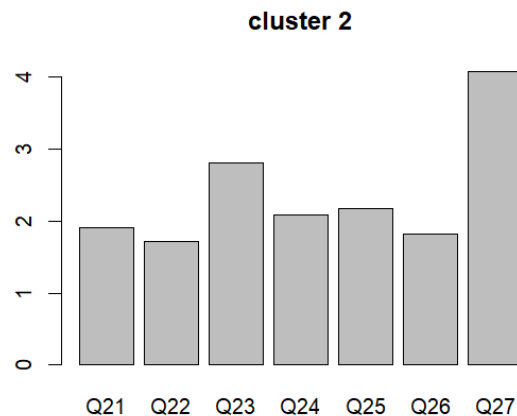
군집분석을 통해 소비가치 5개 집단, 라이프스타일 5개 집단, 모바일 소비 이용행태 4개, 정보활용역량 4개로 범주를 도출함. 이를 이용해 잠재계층분석을 실시하고자 함

차원	항목 구분	정의
소비가치 (5개)	다가치추구형 극단적경제추구형 환경적가치추구형 실용적 소비추구형 감성적 경제추구형	▶ 경제적/환경적/사회적 가치 모두 두루두루 중요시하는 소비자형 ▶ 소비 시 경제적 가치만을 극단적으로 중요시하는 소비자형 ▶ 소비가 미치는 환경적 영향을 중요시하는 소비자형 ▶ 가격, 지속성 등 실용적 측면을 중요시하는 소비자형 ▶ 상황 및 분위기에 따라 소비하되, 경제적 가치를 중요시하는 소비자형
라이프 스타일 (5개)	내향적 사회활동형 보수적/내향적 활동지향형 적극적 사회활동형 혁신적/내향적 활동지향형 멀티테스킹형	▶ 외모/미용 등에 관심은 많으나 내향적이고, 빠른 변화보다는 점진적 변화를 추구하는 유형 ▶ 변화/유행 및 외모/미용 등에 둔감하고, 사회활동보다는 내부 활동을 선호하는 유형 ▶ 외모/미용 등에 대한 관심은 적으나 새로운 것에 관심이 많고, 적극적인 사교활동을 추구하는 유형 ▶ 내향적이지만 혁신적인 변화나 유행에 대한 수용도는 높아 새로운 것에 대해 기대하는 유형 ▶ 외향적이고, 유행에 민감하며, 적극적인 사교활동을 추구하고, 새로운 것을 적극적으로 수용하는 유형
모바일 소비 이용행태 (4개)	맹목적 온라인 의존형 대중적 온라인 이용형 실용적 온라인 이용형 자기주도적 온라인이용형	▶ 직접 눈으로 확인하지 않아도, 온라인이면 무조건 오프라인보다 선호하는 유형 ▶ 평균적인 수준의 가장 대중적 온라인 이용형이며, 많은 대안 및 가격 비교가 주 이용 목적인 유형 ▶ 대안비교 후 오프라인으로 확인, 최종적으로 온라인으로 최저가를 찾아 구매하는 등의 실용적 이용 유형 ▶ 온라인도, 오프라인도 모두 적절히 이용하나 필요에 의해 선택적으로 온라인을 이용하는 유형
정보 활용 역량 (4개)	선택적 정보활용형 소극적 정보활용형 자기주도적 정보활용형 적극적 정보활용형	▶ 필요한 정보의 소스를 잘 알고 있고, 타인의 의견을 참고는 하나 선택적으로 정보를 취하는 유형 ▶ 필요한 정보 소스는 알고 있으나 정보 과잉으로 적극적으로 정보를 활용하지는 않는 유형 ▶ 타인의 의견, 다양한 정보를 활용하기 보다는 자기가 필요한 정보만 적절히 취해 활용하는 유형 ▶ 다양한 소스를 알고 있고, 타인의 의견, 구매자의 추천 등 다양한 정보를 적극 활용하나 정보피로도를 느낌

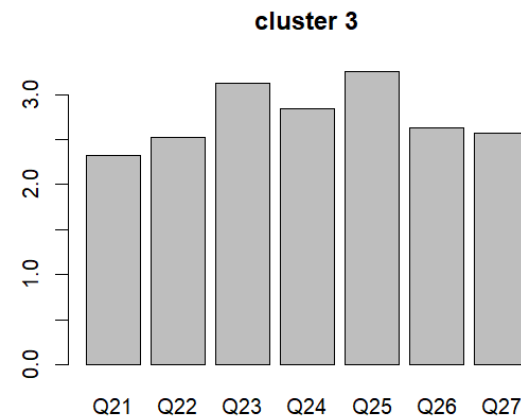
[Back-up] Clustering 결과 - 소비가치



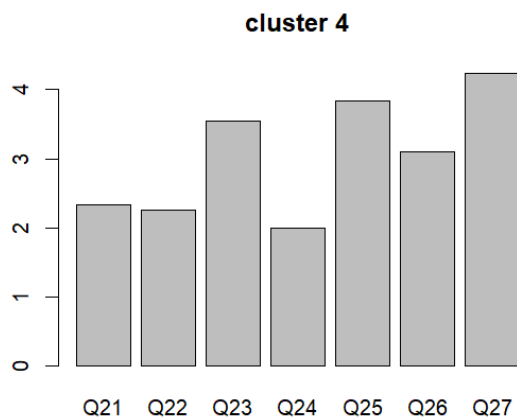
다가치 추구형



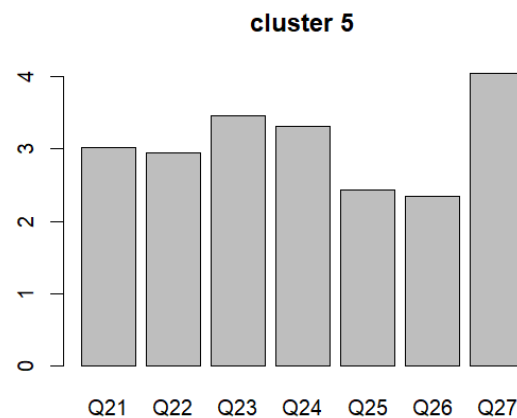
극단적 경제추구형



환경적 가치추구형



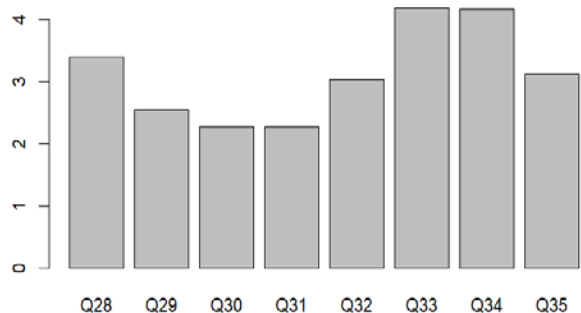
실용적 소비추구형



감성적 경제추구형

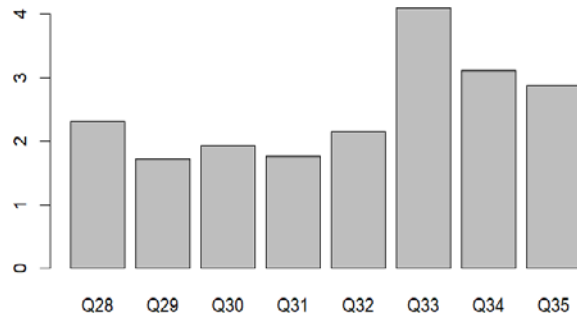
[Back-up] Clustering 결과 - 라이프스타일

cluster 1



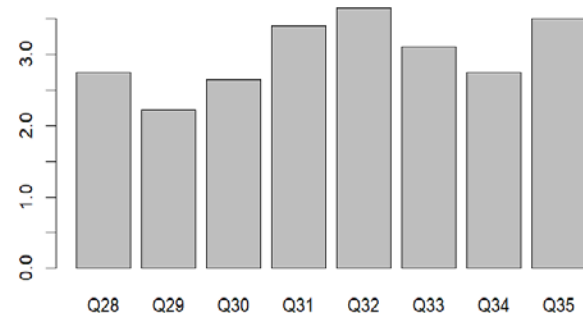
내향적 사회활동형

cluster 2



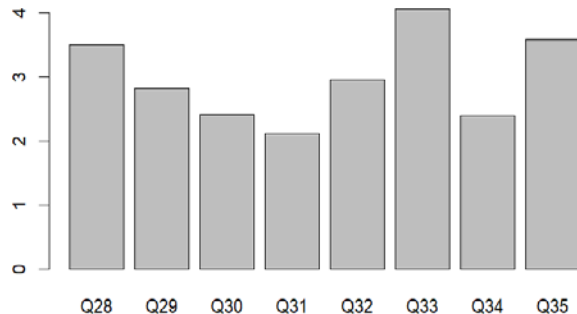
보수적/내향적 라이프스타일

cluster 3



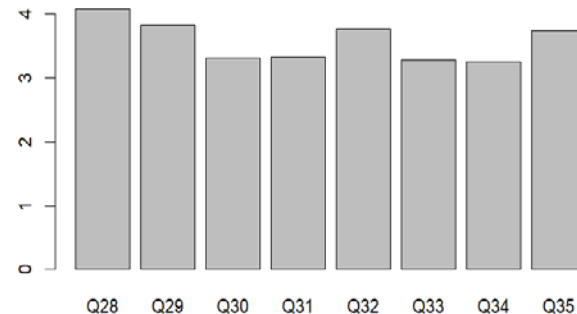
적극적 사회활동형

cluster 4



혁신적/내향적 라이프스타일

cluster 5



멀티테스킹형

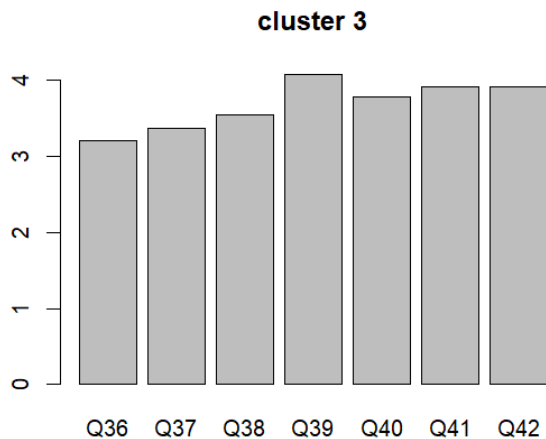
[Back-up] Clustering 결과 - 모바일 이용행태



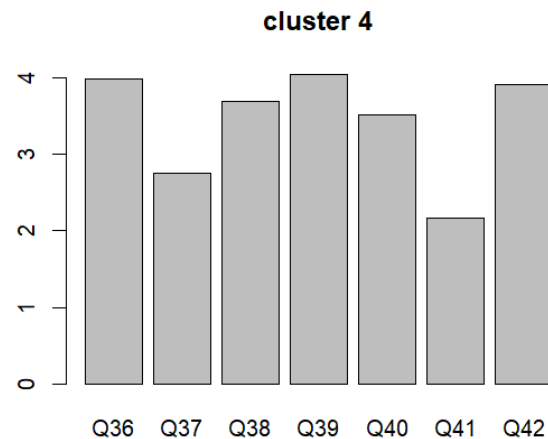
맹목적 온라인 의존형



대중적 온라인 이용형



실용적 온라인 의존형



자기주도적 온라인 이용형

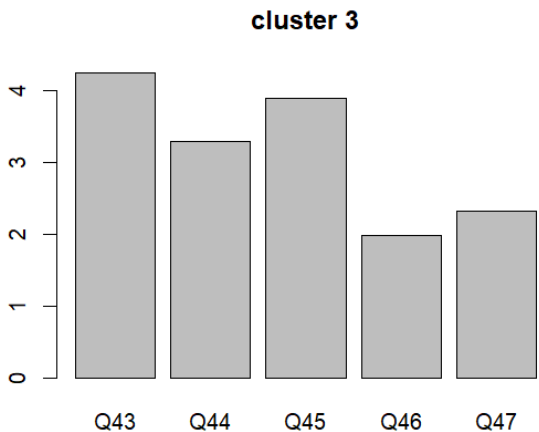
[Back-up] Clustering 결과 - 정보활용 역량



선택적 정보활용형



소극적 정보활용형



자기주도형 정보활용형



적극적 정보활용형

군집분석 결과를 통한 잠재계층분석(LCA) 결과

잠재계층분석(LCA) 결과, 최적 세분시장 개수는 7개*가 최적인 것으로 도출되었으며, 각 시장의 각 항목에 대한 항목-반응 행렬은 아래와 같음

차원	항목	집단별 항목 반응 행렬 (Item-response Matrix)						
		class 1 (0.206)	class 2 (0.08)	class 3 (0.12)	class 4 (0.151)	class 5 (0.097)	class 6 (0.21)	class 7 (0.136)
소비가치	다가치추구형	0.428	0.138	0.651	0.130	0.000	0.101	0.036
	극단적경제추구형	0.065	0.358	0.000	0.000	0.479	0.144	0.299
	환경적가치추구형	0.127	0.436	0.118	0.240	0.080	0.113	0.512
	실용적 소비추구형	0.163	0.068	0.073	0.198	0.250	0.402	0.135
	감성적 경제추구형	0.218	0.000	0.159	0.431	0.191	0.240	0.018
라이프 스타일	내향적 사회활동형	0.000	0.339	0.163	0.248	0.000	0.488	0.000
	보수적/내향적 활동지향형	0.000	0.488	0.000	0.099	0.589	0.177	0.433
	적극적 사회활동형	0.166	0.128	0.102	0.123	0.158	0.185	0.359
	혁신적/내향적 활동지향형	0.308	0.045	0.110	0.439	0.253	0.150	0.122
	멀티테스킹형	0.526	0.000	0.626	0.090	0.000	0.000	0.085
모바일 소비 이용행태	맹목적 온라인 의존형	0.345	0.316	0.090	0.216	0.339	0.373	0.000
	대중적 온라인 이용형	0.047	0.055	0.225	0.269	0.052	0.163	0.800
	실용적 온라인 의존형	0.241	0.499	0.583	0.235	0.159	0.285	0.000
	자기주도적 온라인이용형	0.366	0.131	0.103	0.281	0.450	0.179	0.200
정보활용 역량	선택적 정보활용형	0.277	0.001	0.191	0.000	0.281	0.566	0.097
	소극적 정보활용형	0.000	0.667	0.176	0.711	0.000	0.000	0.618
	자기주도적 정보활용형	0.502	0.000	0.000	0.192	0.644	0.215	0.166
	적극적 정보활용형	0.221	0.332	0.634	0.097	0.075	0.219	0.119

세분 시장 프로파일링 : 집단#1

집단 1은 다양한 가치, 새로움을 적극적으로 수용하고, 적극적인 사교활동을 추구하며 온라인을 적극적으로 활용하며 정보를 자기주도적으로 받아들이는 집단으로 다음과 같은 특성을 지님

집단#1 (n=257, 23.3%)

소비가치

다가치 추구형

라이프스타일

멀티태스킹형

모바일 이용행태

자기주도형/
맹목적 의존형

정보활용 역량

자기주도적 정보활용

연령

30~39세(40.5%)

20 ~ 29세(38.5%)
40 ~ 49세(14.0%)

성별

여성(76.3%)

남성(23.7%)

소득

250~300만원(20.2%)

200~250만원(19.5%)
150 ~ 200만원(13.6%)

학력

대졸 or 재학
(70.4%)

대학원 졸업 or 재학
(20.6%)

직업

회사원 (51.4%)

전문직 (17.9%)

1인 가구 평균기간

6년 11개월

주요 지출비중

외식(22.2%)

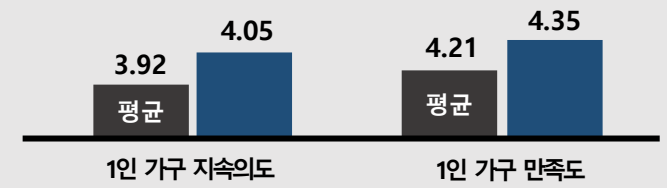
금융활동(18.6%)
주거비(16.9%)
생필품(14.4%)
개인 서비스(3.07%)

1인 가구 사유

학교/직장(48.6%)

혼자 살고 싶어서(21.0%)
미혼이라서(20.6%)

1인 가구 의도 및 만족도



세분 시장 프로파일링 : 집단#2

집단 2는 환경적 가치를 추구하는 소비를 하며, 보수적/내향적인 활동을 주로 추구하고 실용적인 온라인 소비와 소극적인 정보를 활용하는 집단으로 다음과 같은 특성을 지님

집단#2 (n=85, 7.7%)

소비가치

환경적 가치
추구형

라이프스타일

보수적/내향적 활동
및 내향적 활동

모바일 이용행태

실용적 온라인
의존형

정보활용 역량

소극적 정보활용형

연령

30~39세(43.5%)

20 ~ 29세(23.5%)
40 ~ 49세(23.5%)

성별

여성(64.7%)

남성(35.3%)

소득

150~200만원(27.1%)

200~250만원(21.2%)
250 ~ 300만원(18.8%)

학력

대졸 or 재학
(75.3%)

대학원 졸업 or 재학
(8.2%)

직업

회사원 (56.5%)

서비스직 (9.4%)
자영업 (9.4%)

1인 가구 평균기간

8년 1개월

주요 지출비중

외식(21.5%)

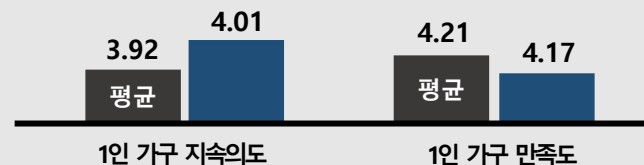
금융활동(20.8%)
생필품(17.1%)
주거비(15.9%)
개인 서비스(2.61%)

1인 가구 사유

미혼이라서(40.0%)

학교/직장 때문(27.1%)
혼자살고 싶어서(24.7%)

1인 가구 의도 및 만족도



세분 시장 프로파일링 : 집단#3

집단 3은 다양한 가치를 추구하는 소비자로 새로움, 유행, 적극적 사교활동 등을 선호하며 실용적인 온라인 이용과 적극적 정보 활용형에 해당하는 소비자로 다음과 같은 특성을 지님

집단#3 (n=123, 11.1%)

소비가치

다가치 추구형

라이프스타일

멀티테스킹형

모바일 이용행태

실용적 온라인의존형

정보활용 역량

적극적 정보활용형

연령

30~39세(43.1%)

20 ~ 29세(34.5%)
50 ~ 59세(13.5%)

성별

여성(60.2%)

남성(39.8%)

소득

250~300만원(22.8%)

200~250만원(22.0%)
150 ~ 200만원(13.0%)
300 ~ 350만원(10.6%)

학력

대졸 or 재학
(70.7%)

대학원 졸업 or 재학
(19.5%)

직업

회사원 (53.6%)

전문직 (10.6%)

1인 가구 평균기간

7년 4개월

주요 지출비중

외식(21.7%)

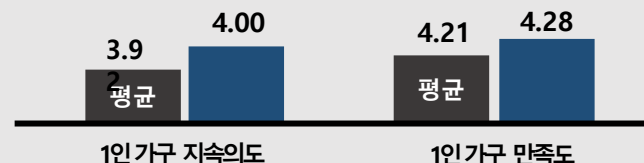
주거비(17.0%)
생필품(16.0%)
금융활동(15.4%)
개인 서비스(3.16%)

1인 가구 사유

학교/직장(39.0%)

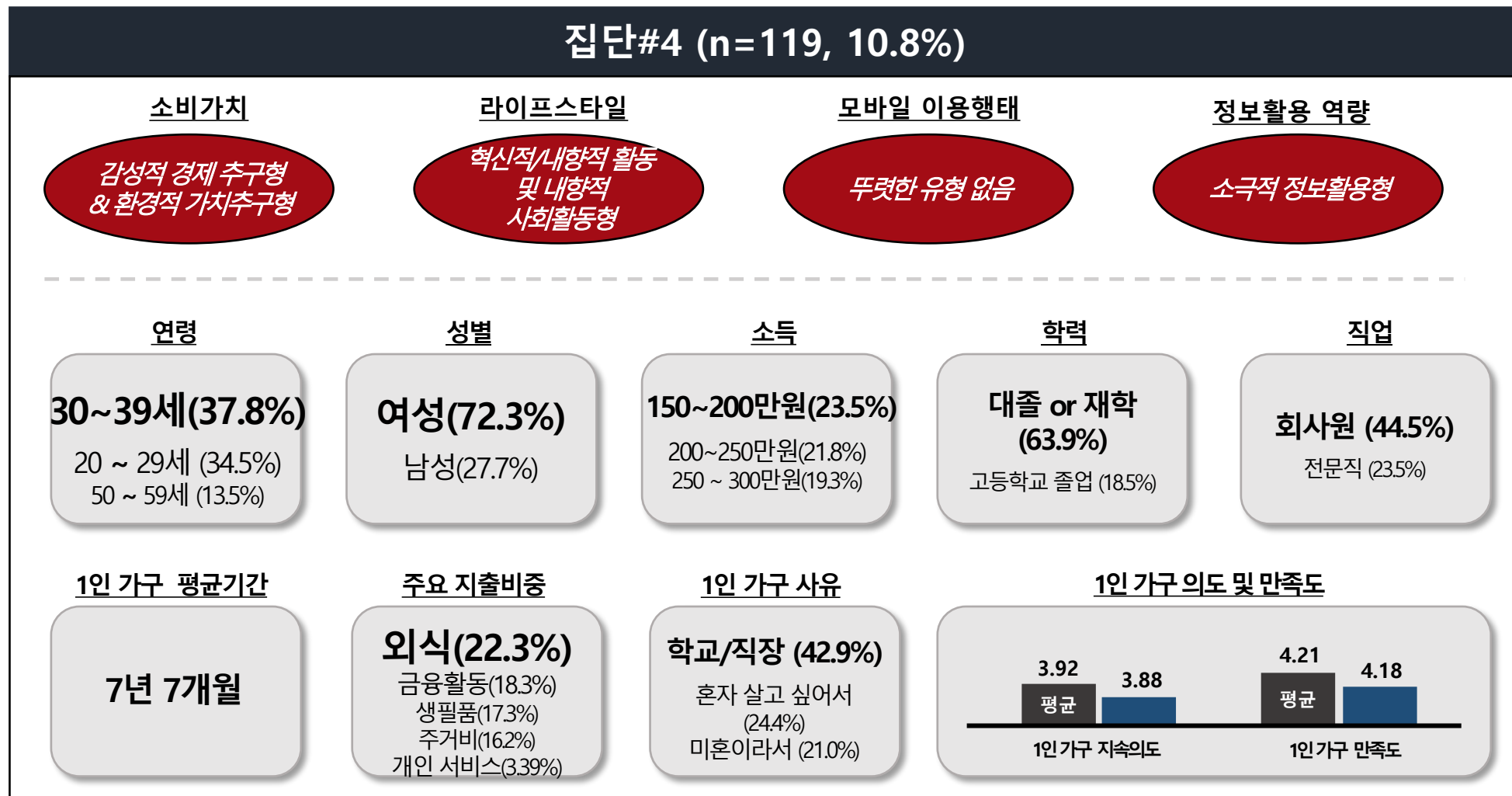
미혼이라서(30.1%)
혼자 살고싶어서(22.8%)

1인 가구 의도 및 만족도



세분 시장 프로파일링 : 집단#4

집단 4는 감성적 경제추구형으로 새로움을 추구하나 외향적 사교활동은 지양하고, 소극적인 정보활용을 나타내는 집단으로 다음과 같은 특성을 지님



세분 시장 프로파일링 : 집단#5

집단 5는 경제적 가치를 극단적으로 추구하는 소비형태를 나타내며, 보수적/내향적인 사회활동을 선호하고, 자기주도적 및 맹목적 온라인 이용과 자기주도적 정보활용을 나타내는 집단으로 다음과 같은 특성을 지님

집단#5 (n=97, 8.8%)

소비가치

극단적 경제추구형
& 실용적 소비추구형

라이프스타일

보수적/내향적
활동지향형

모바일 이용행태

자기주도적
온라인이용형
& 맹목적 의존형

정보활용 역량

자기주도적
정보활용형

연령

30~39세(42.3%)

20 ~ 29세(28.9%)
40 ~ 49세(19.6%)

성별

여성(61.9%)

남성(38.1%)

소득

200~250만원(23.7%)

150~200만원(18.6%)
300 ~ 350만원(14.4%)

학력

대졸 or 재학
(70.1%)

대학원 졸업 or 재학
(19.6%)

직업

회사원 (47.4%)

전문직 (15.5%)
대학(원)생(11.3%)

1인 가구 평균기간

8년 1개월

주요 지출비중

금융(22.1%)

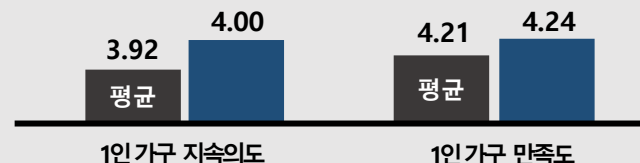
주거비(21.5%)
생필품(18.1%)
외식(16.4%)
개인 서비스(3.16%)

1인 가구 사유

학교/직장(41.2%)

혼자 살고 싶어서(26.8%)
미혼이라서(24.7%)

1인 가구 의도 및 만족도



세분 시장 프로파일링 : 집단#6

집단 6은 실용적 소비 및 감성적 가치를 추구하는 소비자로 내향적인 사회활동과 맹목적 및 실용적 모바일 이용행태를 나타내고, 선택적으로 정보를 활용하는 집단이며 다음과 같은 특성을 지님

집단#6 (n=245, 22.2%)

소비가치

실용적 소비추구형
& 감성적 경제추구형

라이프스타일

내향적 사회활동형

모바일 이용행태

맹목적 & 실용적
온라인 의존형

정보활용 역량

선택적 정보활용형

연령

20~29세(41.2%)

30 ~ 39세(40.4%)
40 ~ 49세(11.4%)

성별

여성(73.5%)

남성(26.5%)

소득

200~250만원(22.0%)

150~200만원(18.0%)
250 ~ 300만원(16.7%)

학력

대졸 or 재학
(69.0%)

대학원 졸업 or 재학
(17.6%)

직업

회사원 (45.3%)

전문직 (13.1%)
서비스직 (11.0%)

1인 가구 평균기간

7년

주요 지출비중

외식(21.5%)

금융활동(20.9%)
생필품(16.7%)
주거비(16.6%)
개인 서비스(3.69%)

1인 가구 사유

학교/직장 (49.0%)

미혼이라서(20.8%)
혼자 살고 싶어서(20.4%)

1인 가구 의도 및 만족도

3.92

평균

3.91

1인 가구 지속의도

4.21

평균

4.21

1인 가구 만족도

세분 시장 프로파일링 : 집단#7

집단 7은 환경적 가치를 추구하는 집단으로 보수적/내향적이나 적극적 활동형도 포함된 집단으로, 대중적인 온라인 이용과 소극적 정보활용 행태를 나타내는 집단으로 다음과 같은 특성을 지님

집단#7 (n=178, 16.1%)

소비가치

환경적 가치 추구형

라이프스타일

보수적/내향적
활동지향형 &
적극적 사회활동형

모바일 이용행태

대중적 온라인
이용형

정보활용 역량

소극적 정보활용형

연령

30~39세(36.0%)
20 ~ 29세(23.0%)
50 ~ 59세(19.6%)

성별

여성(57.3%)
남성(42.7%)

소득

150~200만원(21.3%)
200~250만원(19.1%)
250 ~ 300만원(19.1%)

학력

대졸 or 재학 (64.6%)
고등학교 졸업 (19.7%)

직업

회사원 (51.7%)
전문직 (13.5%)

1인 가구 평균기간

8년 11개월

주요 지출비중

금융(21.4%)
외식(18.8%)
주거비(17.6%)
생필품(17.1%)
개인 서비스(3.61%)

1인 가구 사유

학교/직장 (34.3%)
미혼이라서 (32.6%)
혼자 살고 싶어서 (16.3%)

1인 가구 의도 및 만족도

