2021 Spring

Artificial Intelligence & Deep Learning

Prof. Minsuk Koo

Department of Computer Science & Engineering
Incheon National University

6.1 지도 학습과 비지도 학습, 준지도 학습

■ 세 가지 유형의 학습

- 지도 학습: 모든 훈련 샘플이 레이블 정보를 가짐
- 비지도 학습: 모든 훈련 샘플이 레이블 정보를 가지지 않음 ← 6장의 주제
- 준지도 학습: 레이블을 가진 샘플과 가지지 않은 샘플이 섞여 있음 ← 7장의 주제

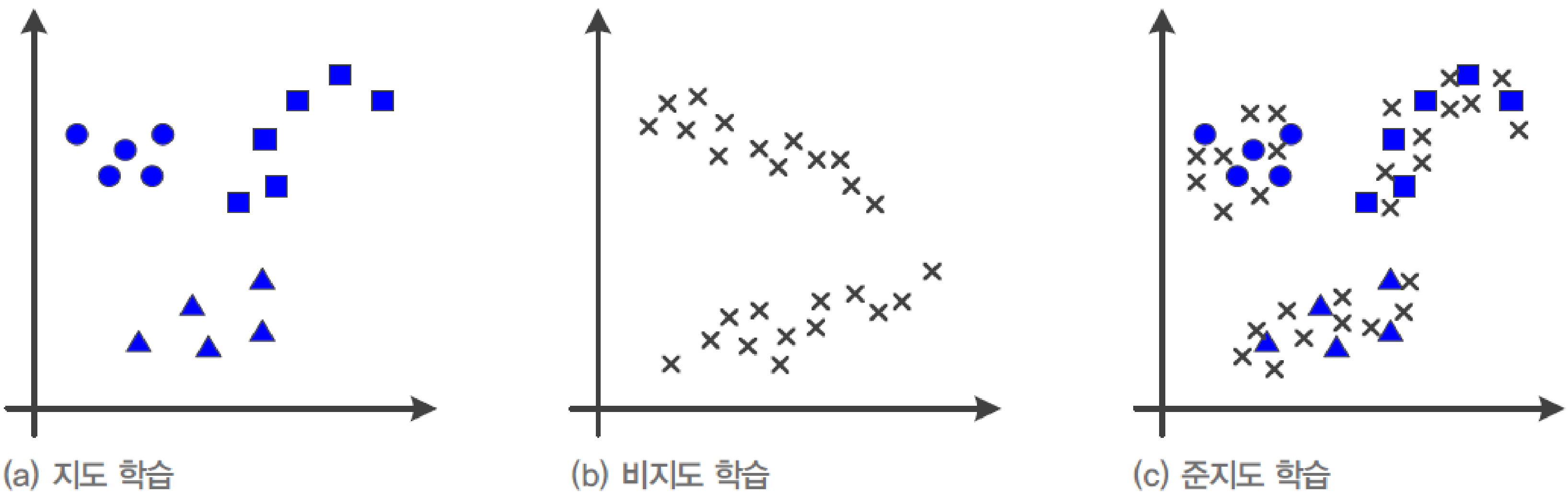


그림 6-1 기계 학습의 유형(속이 찬 샘플은 레이블이 있고, x 표시된 샘플은 레이블이 없음)

6.1 지도 학습과 비지도 학습, 준지도 학습

- 기계 학습이 사용하는 두 종류의 지식
  - 훈련집합
  - 사전 지식prior knowledge(세상의 일반적인 규칙)

데이터가
낮은 차원으로 표현할수있다. 라는가정

- 중요한 두 가지 사전 지식
  - 매니폴드 가정manifold hypothesis: 데이터집합은 하나의 매니폴드 또는 여러 개의 매니폴드를 구성하며, 모든 샘플은 매니폴드와 가까운 곳에 있다. 매니폴드와 매니폴드 가정은 6.8.1절에서 자세히 설명한다.

급작스럽게 변하지 않는다.

  - 매끄러움 가정smoothness hypothesis: 샘플은 어떤 요인에 의해 변화한다. 예를 들어, 장면과 카메라 위치를 고정한 상태에서 조명을 조금씩 변화하면서 영상을 획득한 경우, 획득된 영상 샘플은 특징 공간에서 위치가 조금씩 바뀔 것이다. 이때 [그림 6-1(b)]와 같이 매끄러운 곡면을 따라 위치가 변한다.
- 비지도 학습과 준지도 학습은 사전 지식을 더 명시적으로 사용

6.2 비지도 학습

- 6.2.1 비지도 학습의 일반 과업
- 6.2.2 비지도 학습의 응용 과업

6.2.1 비지도 학습의 일반 과업

- 세 가지 일반 과업
  - 군집화: 유사한 샘플을 모아 같은 그룹으로 묶는 일
  - 밀도 추정: 데이터로부터 확률분포를 추정하는 일
  - 공간 변환: 원래 특징 공간을 저차원 또는 고차원 공간으로 변환하는 일
- 데이터에 내재한 구조를 잘 파악하여 새로운 정보를 발견해야 함

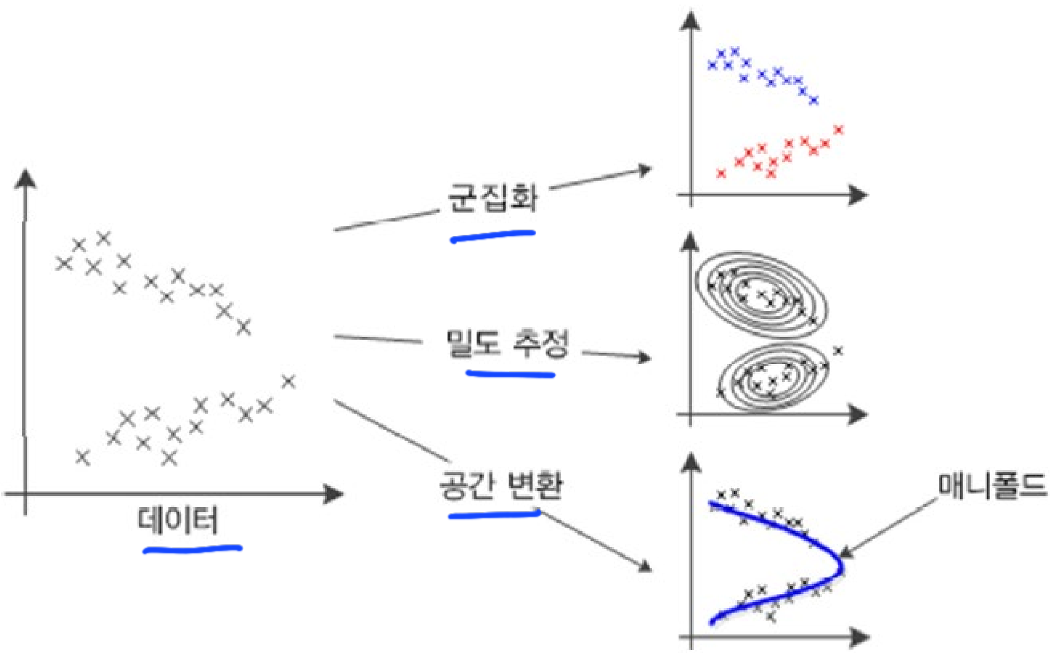


그림 6-2 비지도 학습의 군집화, 밀도 추정, 공간 변환 과업이 발견하는 정보

6.2.2 비지도 학습의 응용 과업

- 아주 많은 응용(서로 밀접하게 연관)
  - 군집화의 응용
    - 맞춤 광고, 영상 분할, 유전자 데이터 분석, SNS 실시간 검색어 분석하여 사람들의 관심 파악 등
  - 밀도 추정의 응용
    - 분류, 생성 모델 구축 등
  - 공간 변환의 응용
    - 데이터 가시화, 데이터 압축, 특징 추출(표현 학습) 등

6.3 군집화

- 6.3.1 *k*-평균 알고리즘 k-means
- 6.3.2 친밀도 전파 알고리즘

6.3 군집화

■ 군집화 문제

- $\mathbb{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$에서 식 (6.1)을 만족하는 군집집합 $C = \{c_1, c_2, \cdots, c_k\}$를 찾아내는 작업

$$\left.\begin{array}{l} c_i \neq \emptyset, i = 1,2,\cdots,k \\ \bigcup_{i=1}^{k} c_i = \mathbb{X} \\ c_i \bigcap c_j = \emptyset, i \neq j \end{array}\right\} \qquad (6.1)$$

- 군집의 개수 k는 주어지는 경우와 자동으로 찾아야 하는 경우가 있음
- 군집화를 부류 발견 작업이라 부르기도 함

k-means : k정해줘야 함
친밀도 : k가 주어짐

■ 군집화의 주관성

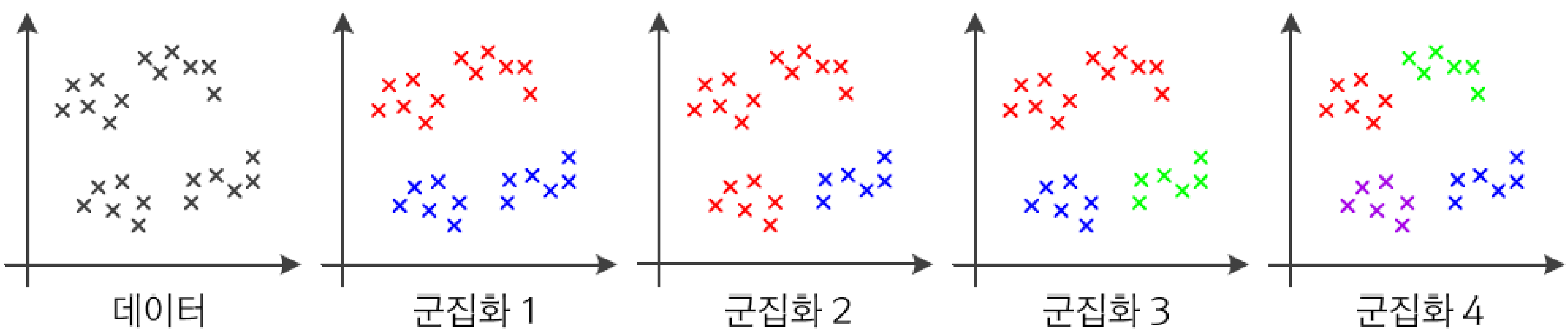


그림 6-3 군집화의 주관성

6.3.1 k-평균 알고리즘

k-평균 알고리즘의 특성

- 원리 단순하지만 성능이 좋아 인기 좋음
- 직관적으로 이해하기 쉽고 구현 쉬움
- 군집 개수 k를 알려줘야 함

알고리즘 6-1 k-평균

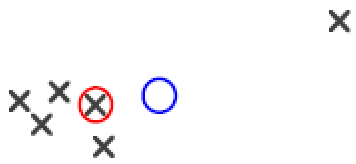
입력: 훈련집합 $\mathbb{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$, 군집의 개수 k
출력: 군집집합 $C = \{c_1, c_2, \cdots, c_k\}$

```
1  k개의 군집 중심 Z = {z_1, z_2, …, z_k}를 초기화한다.
2  while (true)
3     for (i=1 to n)
4        x_i를 가장 가까운 군집 중심에 배정한다.   ~> 유클리디안 거리
5     if (라인 3~4에서 이루어진 배정이 이전 루프에서의 배정과 같으면) break
6     for (j=1 to k)
7        z_j에 배정된 샘플의 평균으로 z_j를 대치한다.   ~> 새로운 대표값
8  for (j=1 to k)
9     z_j에 배정된 샘플을 c_j에 대입한다.
```

6.3.1 *k*-평균 알고리즘

k-평균과 *k*-medoids

- *k*-평균은 [알고리즘 6-1]의 라인 7에서 샘플의 평균으로 군집 중심을 갱신
- *k*-medoids는 대표를 뽑아 뽑힌 대표로 군집 중심을 갱신(*k*-평균에 비해 잡음에 둔감)



○ *k*-평균에 의한 새로운 군집 중심 ~> 실제값에 없을 수도

○ *k*-medoids에 의한 새로운 군집 중심 ~> 실제값 중 대표값

그림 6-4 *k*-평균과 *k*-medoids가 군집 중심을 갱신하는 과정

최적화 문제로 해석

- *k*-평균은 식 (6.2)의 목적함수를 최소화하는 알고리즘
- 행렬 $\mathbf{A}$는 군집 배정 정보를 나타내는 $k*n$ 행렬(i번째 샘플이 j번째 군집에 배정되었다면 a_{ji}는 1, 그렇지 않으면 0)

$$J(Z,\mathbf{A}) = \sum_{i=1}^{n}\sum_{j=1}^{k} a_{ji}\, dist(\mathbf{x}_i, \mathbf{z}_j) \qquad (6.2)$$

└> 데이터가 어디에 포함되어 있는지 (one hot code)

예제 6-1 ***k*-평균의 동작**

[그림 6-5]는 훈련집합이 7개의 샘플을 가진 n=7인 예를 보여 준다. 좌표는 다음과 같다.

$$\mathbf{x}_1 = \begin{pmatrix} 18 \\ 5 \end{pmatrix},\ \mathbf{x}_2 = \begin{pmatrix} 20 \\ 9 \end{pmatrix},\ \mathbf{x}_3 = \begin{pmatrix} 20 \\ 14 \end{pmatrix},\ \mathbf{x}_4 = \begin{pmatrix} 20 \\ 17 \end{pmatrix},\ \mathbf{x}_5 = \begin{pmatrix} 5 \\ 15 \end{pmatrix},\ \mathbf{x}_6 = \begin{pmatrix} 9 \\ 15 \end{pmatrix},\ \mathbf{x}_7 = \begin{pmatrix} 6 \\ 20 \end{pmatrix}$$

군집의 개수 k=3이라 하자. 맨 왼쪽 그림은 초기 군집 중심을 보여 준다. [알고리즘 6-1]의 라인 3~4는 7개 샘플을 아래와 같이 배정할 것이다.

$$\{\mathbf{x}_1\}\text{은 } \mathbf{z}_1,\ \{\mathbf{x}_2\}\text{은 } \mathbf{z}_2,\ \{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7\}\text{은 } \mathbf{z}_3$$

이 배정을 행렬 **A**로 표현하면 다음과 같다.

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

6.3.1 *k*-평균 알고리즘

[그림 6-5]의 가운데 그림은 새로 계산한 군집 중심이다. $\mathbf{z}_1 = (18, 5)^{\mathrm{T}}$, $\mathbf{z}_2 = (20, 9)^{\mathrm{T}}$, $\mathbf{z}_3 = (12, 16.2)^{\mathrm{T}}$이고, 식 (6.2)에 대입하면 $J =$ 33.77 이 된다. 이때 거리함수 dist로 식 (1.7)의 유클리디언 거리를 사용한다.

두 번째 루프를 실행하면 행렬 **A**는 아래와 같이 바뀐다. 군집 중심은 $\mathbf{z}_1 = (18, 5)^{\mathrm{T}}$, $\mathbf{z}_2 = (20, 13.333)^{\mathrm{T}}$, $\mathbf{z}_3 = (6.667, 16.667)^{\mathrm{T}}$이다. 이것을 식 (6.2)에 대입하면 J = 17.29 이 된다. [그림 6-5]의 맨 오른쪽 그림은 두 번째 루프 수행 후의 상황이다.

$$\mathbf{A} = \begin{array}{cl} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} & \begin{array}{l} \rightarrow \text{첫번째 군집} \\ \rightarrow \text{두번째 〃} \\ \rightarrow \text{세번째 〃} \end{array} \end{array}$$

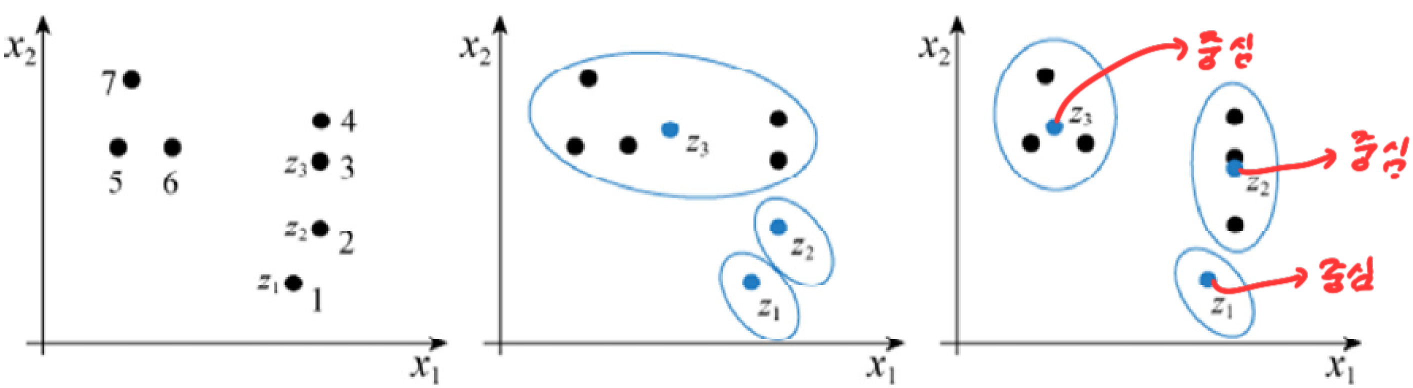


그림 6-5 *k*-평균의 동작 예제

6.3.1 *k*-평균 알고리즘

다중 시작 *k*-평균

- *k*-평균은 [알고리즘 6-1]의 라인 1에서 초기 군집 중심이 달라지면 최종 결과가 달라짐
- 다중 시작은 서로 다른 초기 군집 중심을 가지고 여러 번 수행한 다음, 가장 좋은 품질의 해를 취함

알고리즘 6-2 다중 시작 *k*-평균

입력: 훈련집합 $\mathbb{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$, 군집의 개수 k, 다중 시작 횟수 t
출력: 군집집합 $C = \{c_1, c_2, \cdots, c_k\}$

```
1  for (i=1 to t)
2      X에서 임의로 k개 샘플을 뽑는다.
3      라인 2에서 뽑은 샘플을 초기 군집 중심으로 삼고, [알고리즘 6-1]의 k-평균을 수행한다.
4      k-평균이 출력한 해를 가지고 식 (6.2)의 목적함숫값을 계산한다.
5  t개의 해 중 목적함숫값이 가장 작은 해를 최종해로 취한다.
```

6.3.1 *k*-평균 알고리즘

EM 기초

- *k*-평균에서 훈련집합 $\mathbb{X}$와 군집집합 C(행렬 $\mathbf{A}$)는 각각 입력단과 출력단에서 관찰 가능
- 중간 단계의 임시 변수 Z(입출력단에서 보이지 않기 때문에 은닉변수라latent variable 부름)
- *k*-평균은 Z의 추정(E 단계)과 $\mathbf{A}$의 추정(M 단계)을 번갈아 가면 수행하는 EM 알고리즘

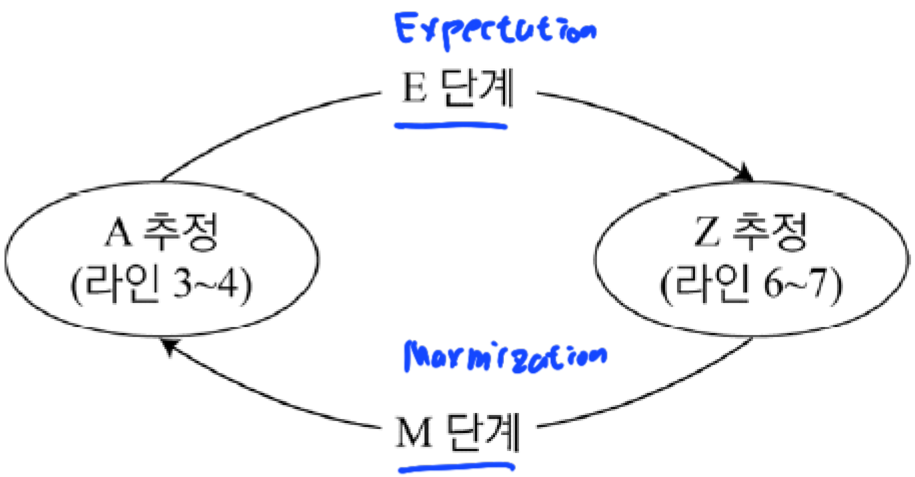


그림 6-6 *k*-평균을 EM 알고리즘으로 해석

6.3.2 친밀도 전파 알고리즘

- 친밀도 전파 알고리즘
  - 책임 행렬 $\mathbf{R}$과 가용 행렬 $\mathbf{A}$라는 두 종류의 친밀도 행렬을 이용하여 군집화
  - 군집 개수 k를 자동으로 알아냄

- 샘플 i와 k의 유사도 s_{ik}

$$s_{ik} = -\|\mathbf{x}_i - \mathbf{x}_k\|_2^2, \qquad i \neq k\text{이고 } i,k = 1,2,\cdots,n \tag{6.3}$$

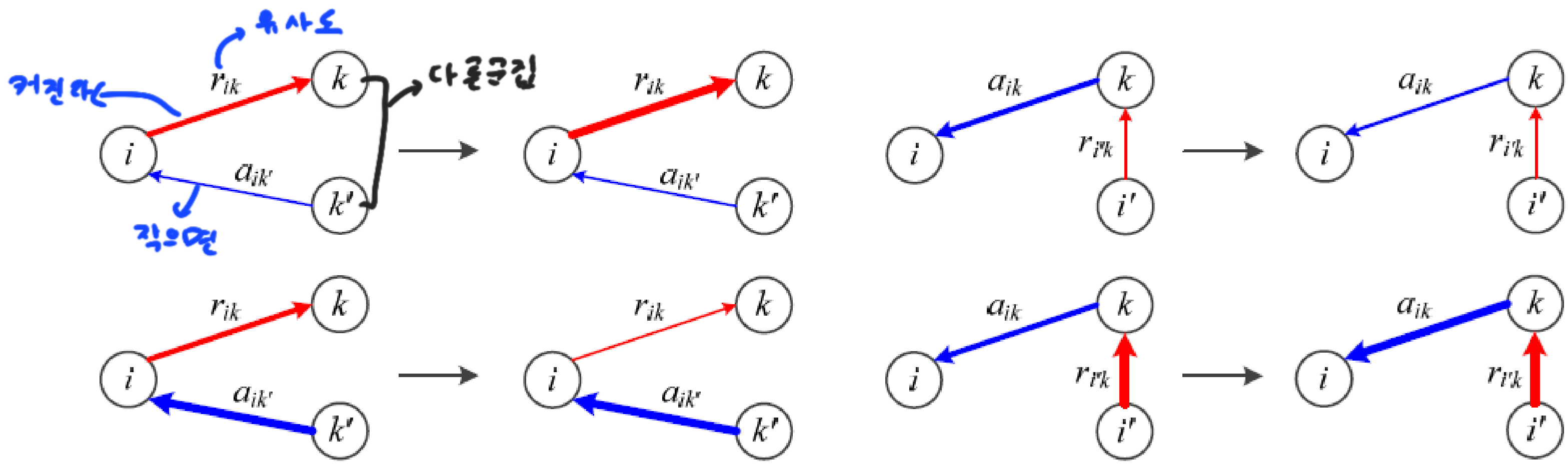
- Euclidian distance

6.3.2 친밀도 전파 알고리즘

- 책임 행렬 **R**과 가용 행렬 **A**의 계산

$$r_{ik} = s_{ik} - \max_{k' \neq k}(a_{ik'} + s_{ik'}) \qquad (6.4)$$

$$a_{ik} = \min\left(0, r_{kk} + \sum_{i' \neq i,k} \max(0, r_{i'k})\right), i \neq k \qquad (6.5)$$



(a) r_{ik}의 계산 (반비례)

(b) a_{ik}의 계산 (비례)

그림 6-7 친밀도 계산 과정

6.3.2 친밀도 전파 알고리즘

- 자가 유사도 s_{kk}
  - 유사도의 최솟값, 중앙값(메디안), 최댓값 중에서 선택(하이퍼 매개변수임)
  - 최솟값은 적은 수의 군집, 최댓값은 많은 수의 군집을 생성. 중앙값은 중간 정도
- 자가 친밀도 r_{kk}와 a_{kk}
  - r_{kk}는 식 (6.4)를 그대로 사용. 즉
  - a_{kk}는 식 (6.6)으로 계산

$$r_{kk} = s_{kk} - \max_{k' \neq k}(a_{kk'} + s_{kk'})$$

$$a_{kk} = \sum_{i' \neq k} \max(0, r_{i'k}) \tag{6.6}$$

6.3.2 친밀도 전파 알고리즘

알고리즘 6-3 친밀도 전파 군집화

s_{kk}

입력: 훈련집합 $\mathbb{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$, 군집 개수 선택사항∈{최솟값, 메디안, 최댓값}, 댐핑 인자 λ

출력: 군집 정보 $\mathbf{z} = (z_1, z_2, \cdots, z_n)^{\mathrm{T}}$

1 for (모든 샘플 쌍 i와 k에 대해) if ($i \neq k$) $s_{ik} = -\|\mathbf{x}_i - \mathbf{x}_k\|_2^2$ // 식 (6.3)

2 for ($k = 1$ to n) s_{kk}=선택 사항에 따라 라인 1의 유사도의 최솟값, 중앙값, 또는 최댓값

3 for (모든 샘플 쌍 i와 k에 대해) $a_{ik}^0 = 0, r_{ik}^0 = 0$

$A = 0 \quad R = 0$

4 $t = 0$

5 repeat

6 $\quad t$++

7 $\quad$ for (모든 샘플 쌍 i와 k에 대해)

8 $\quad\quad r_{ik}^t = s_{ik} - \max_{k' \neq k}\left(a_{ik'}^t + s_{ik'}\right)$ // 식 (6.4)

9 $\quad\quad r_{ik}^t = \lambda r_{ik}^t + (1-\lambda) r_{ik}^{t-1}$

10 $\quad$ for (모든 샘플 쌍 i와 k에 대해)

11 $\quad\quad$ if ($i \neq k$) $a_{ik}^t = \min\left(0, r_{kk}^t + \sum_{i' \neq i,k} \max\left(0, r_{i'k}^t\right)\right)$ // 식 (6.5)

12 $\quad\quad$ else $a_{kk}^t = \sum_{i' \neq k} \max\left(0, r_{i'k}^t\right)$ // 식 (6.6)

13 $\quad\quad a_{ik}^t = \lambda a_{ik}^t + (1-\lambda) a_{ik}^{t-1}$

14 until (멈춤 조건)

15 for ($i = 1$ to n)

16 $\quad \hat{k} = \mathrm{argmax}_{k=1,n}\left(a_{ik}^t + r_{ik}^t\right)$

17 $\quad$ if ($\hat{k} = i$) $z_i = i$ // 이 샘플은 군집 중심임

18 $\quad$ else $z_i = \hat{k}$ // 이 샘플은 $\hat{k}$ 군집 중심에 속함

라인 9&13은 진자 현상을
방지하기 위해 댐핑 인자 적용

← 자신의 자가 친밀도가 최고이면 군집 대표

6.4 밀도 추정

- 6.4.1 커널 밀도 추정
- 6.4.2 가우시안 혼합
- 6.4.3 EM 알고리즘

- 밀도 추정 문제
  - 어떤 점 $\mathbf{x}$에서 데이터가 발생할 확률, 즉 확률 분포 $P(\mathbf{x})$를 구하는 문제
  - 예를 들어, 그림 6-8에서 $P(\mathbf{x}_1) > P(\mathbf{x}_2) > P(\mathbf{x}_3)$

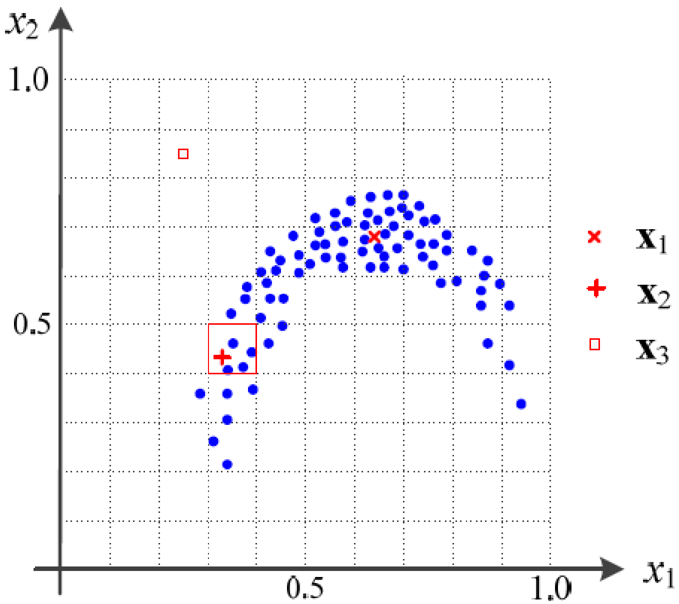


그림 6-8 밀도 추정 문제