

02. 기계학습 (Machine Learning)

- 기계학습의 정의
- 지도학습과 비지도학습
- 신경망과 깊은학습 개요

머신러닝

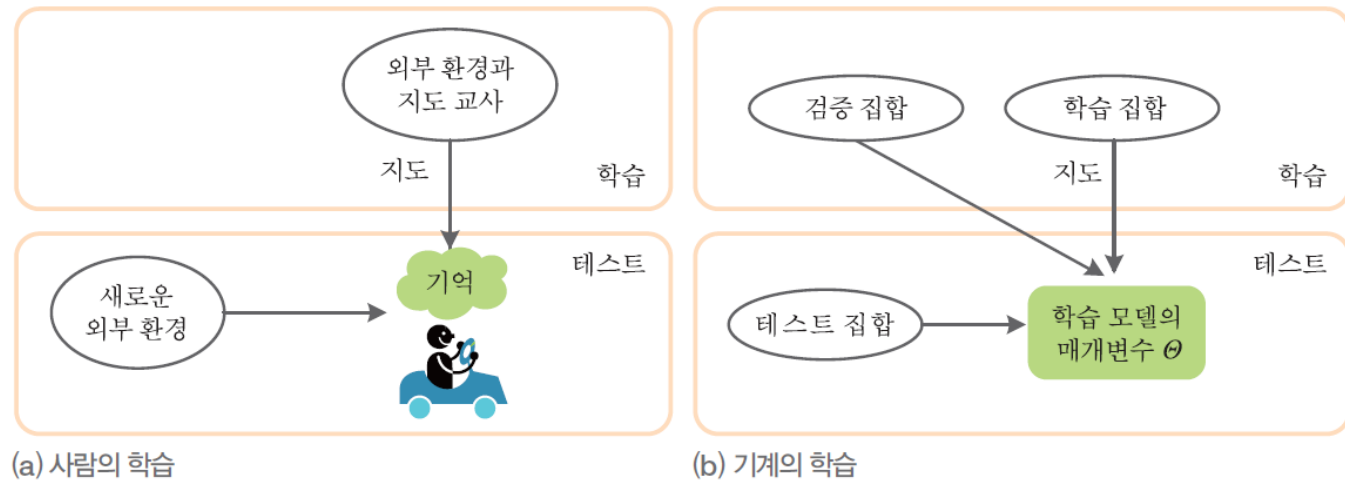
■ Mitchell의 정의

- “A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**.”
- → 어떤 컴퓨터 프로그램이 **T**라는 작업을 수행한다. 이 프로그램의 성능을 **P**라는 척도로 평가했을 때 경험 **E**를 통해 성능이 개선된다면 이 프로그램은 학습을 한다고 말할 수 있다.
 - **T** = 작업(분류)
 - **E** = 경험(학습)
 - **P** = 성능평가척도(인식률)

학습과정

■ 사람과 기계의 학습 과정

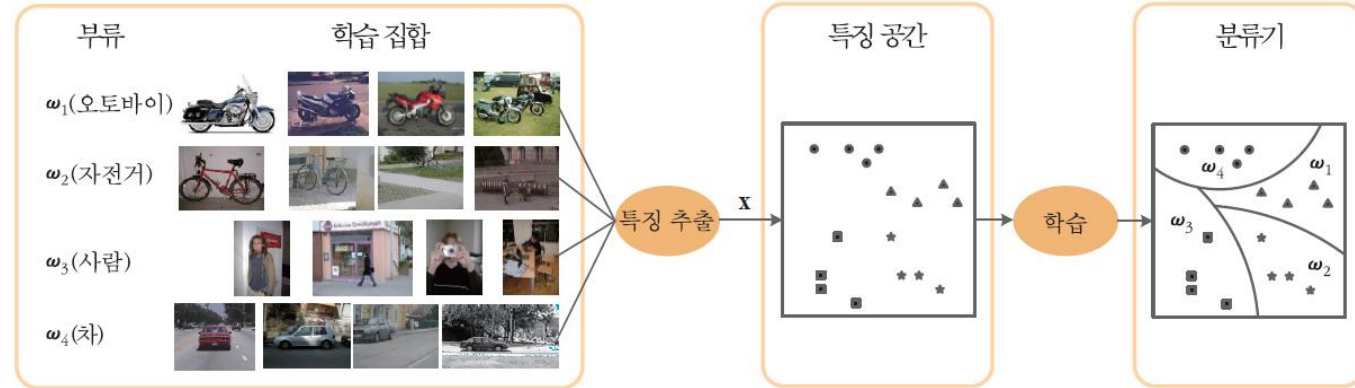
- 학습 단계: 외부 지도에 따라 배우는 과정 (학습 집합 사용)
- 테스트 단계: 현장에서 성능을 평가 하는 과정 (테스트 집합 사용)



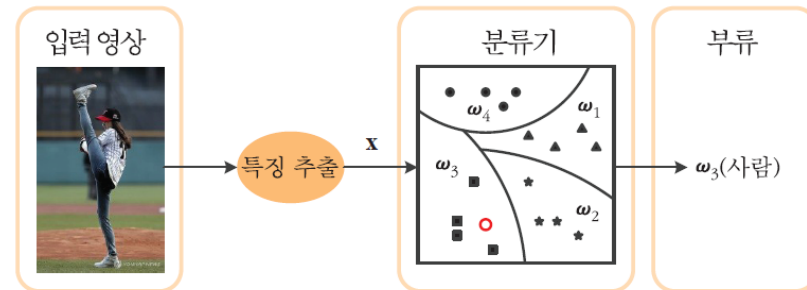
- 지도학습과 비지도학습으로 구분

지도 학습(supervised learning)

- 부류 정보를 가진 샘플 (x, t) 로 구성된 학습 집합 사용
 - x 는 특징 벡터이고 t 는 x 가 속한 부류



(a) 학습 단계



(b) 테스트(인식) 단계

비지도 학습(non-supervised learning)

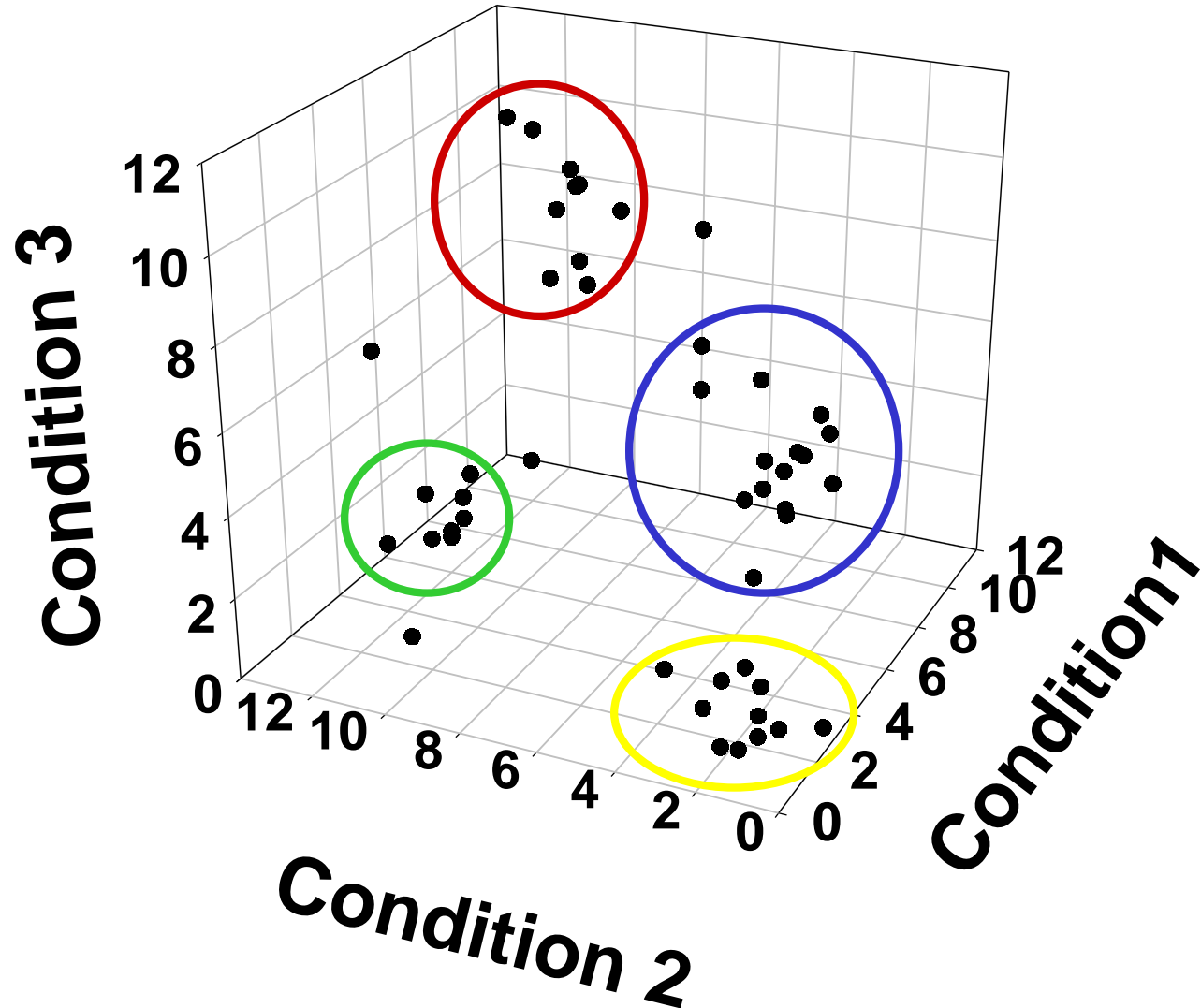
■ 비지도 학습

- 지도 학습에서 사용했던 (x, t) 중에 부류 정보 t 가 없는 상황의 학습
- 유사한 특징 벡터들을 모으는 군집화 수행(K-means, SOM 신경망, 민시프트 등의 군집화 알고리즘)
- 군집에서 유용한 정보 추출(데이터 마이닝, 빅데이터, 정보 검색 등에 응용)

■ 준지도 학습

- 부류 정보가 있는 샘플과 없는 샘플이 섞여 있는 상황의 학습
- 원리
 - 부류 정보가 있는 샘플로 학습한 후 부류 정보가 없는 샘플의 부류 정보를 추정
 - 추정된 정보로 반복 학습

참고: Feature space & Clustering



K-means Algorithm

- For a given assignment C , compute the cluster means m_k :

$$m_k = \frac{\sum_{i:C(i)=k} x_i}{N_k}, k = 1, \dots, K.$$

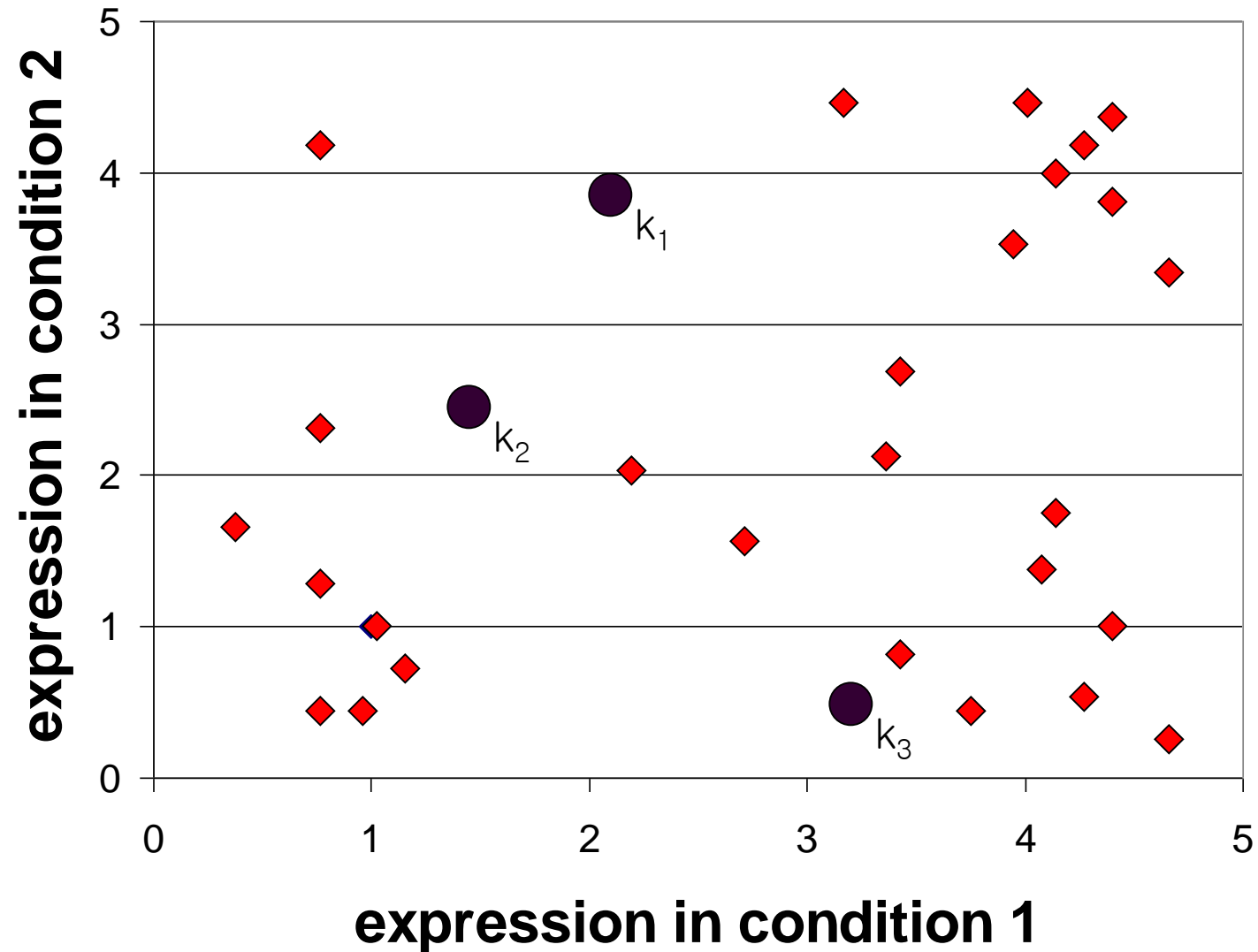
- For a current set of cluster means, assign each observation as:

$$C(i) = \arg \min_{1 \leq k \leq K} \|x_i - m_k\|^2, i = 1, \dots, N$$

- Iterate above two steps until convergence

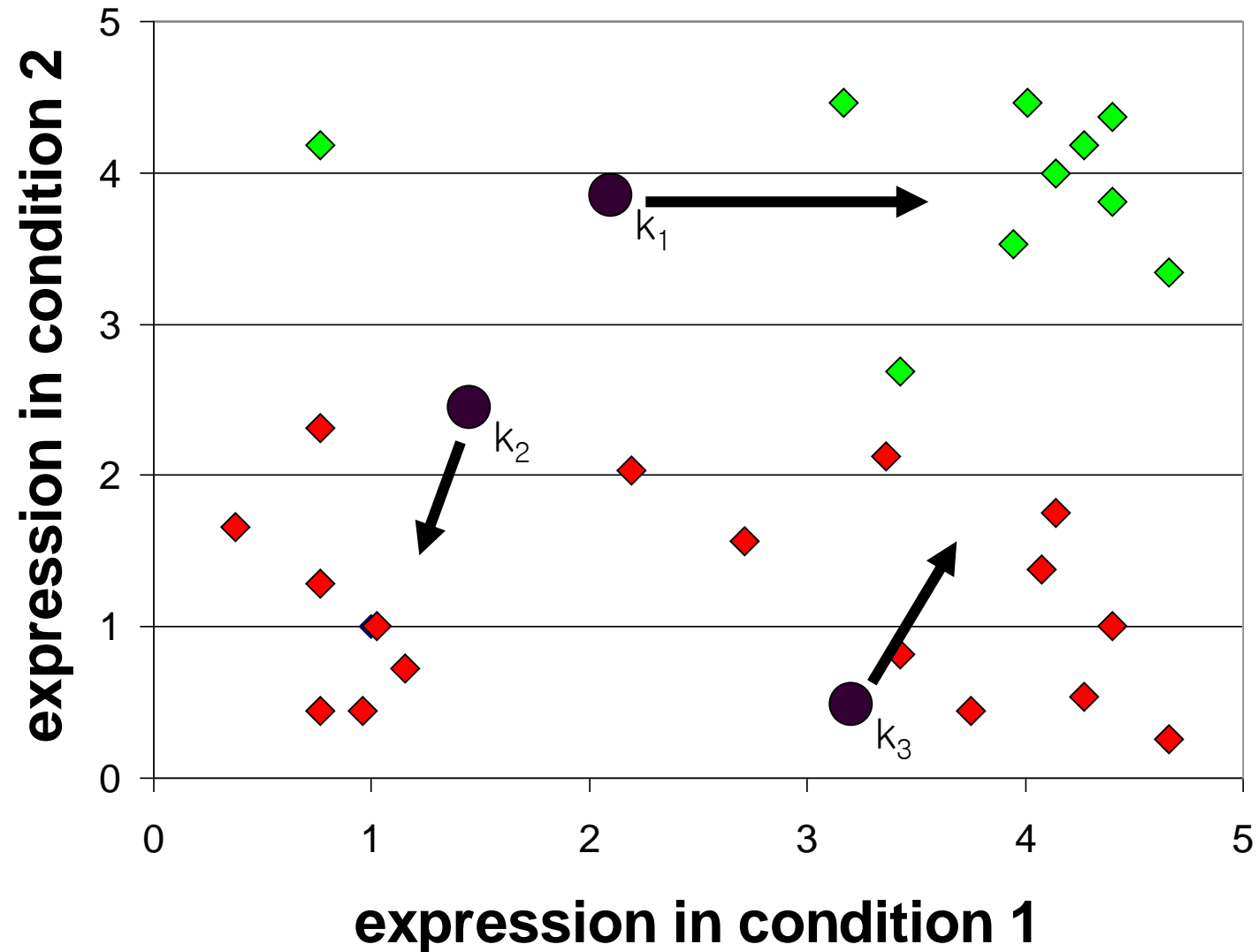
Clustering: Example, Step 1

Algorithm: k-means, Distance Metric: Euclidean Distance



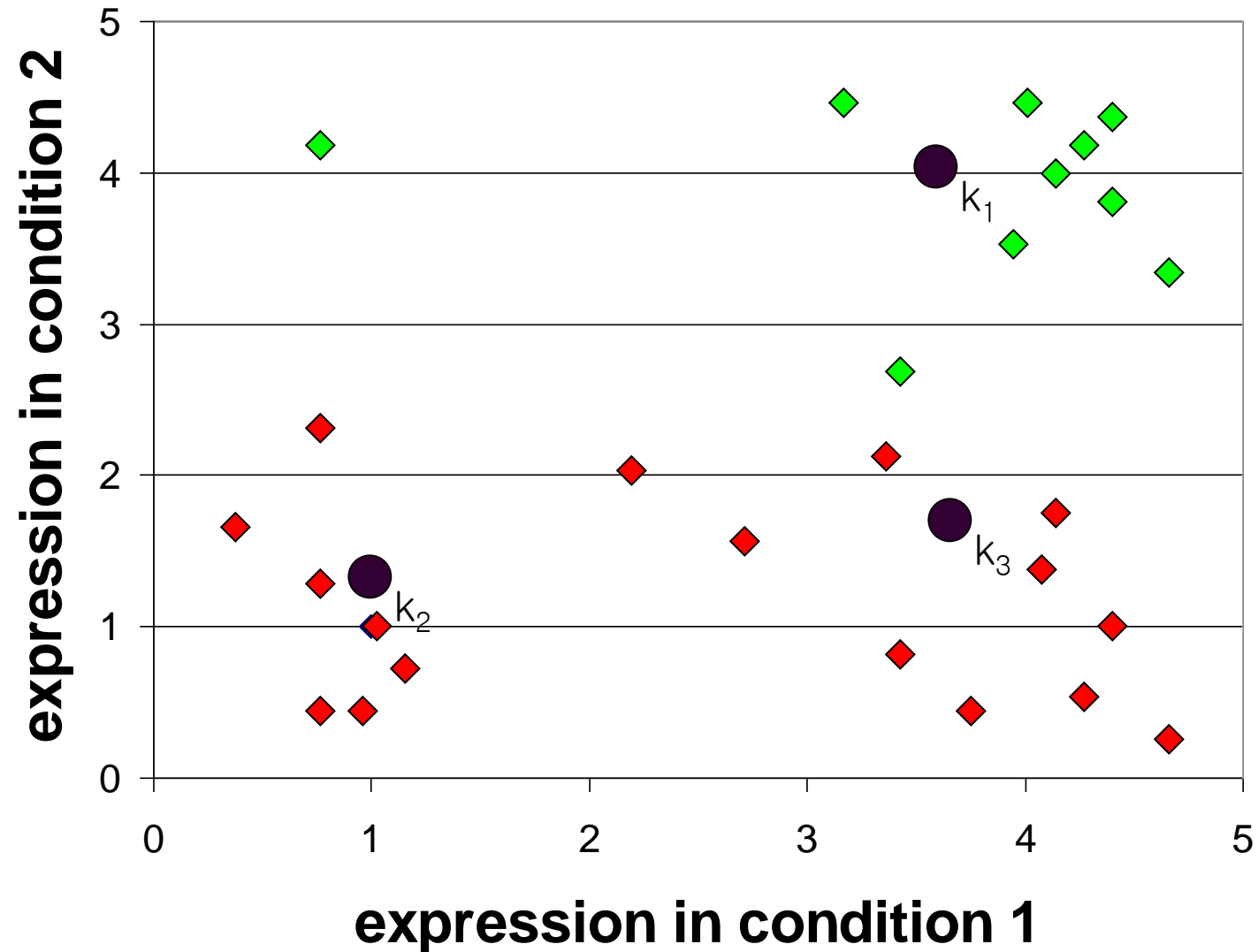
Clustering: Example, Step 2

Algorithm: k-means, Distance Metric: Euclidean Distance



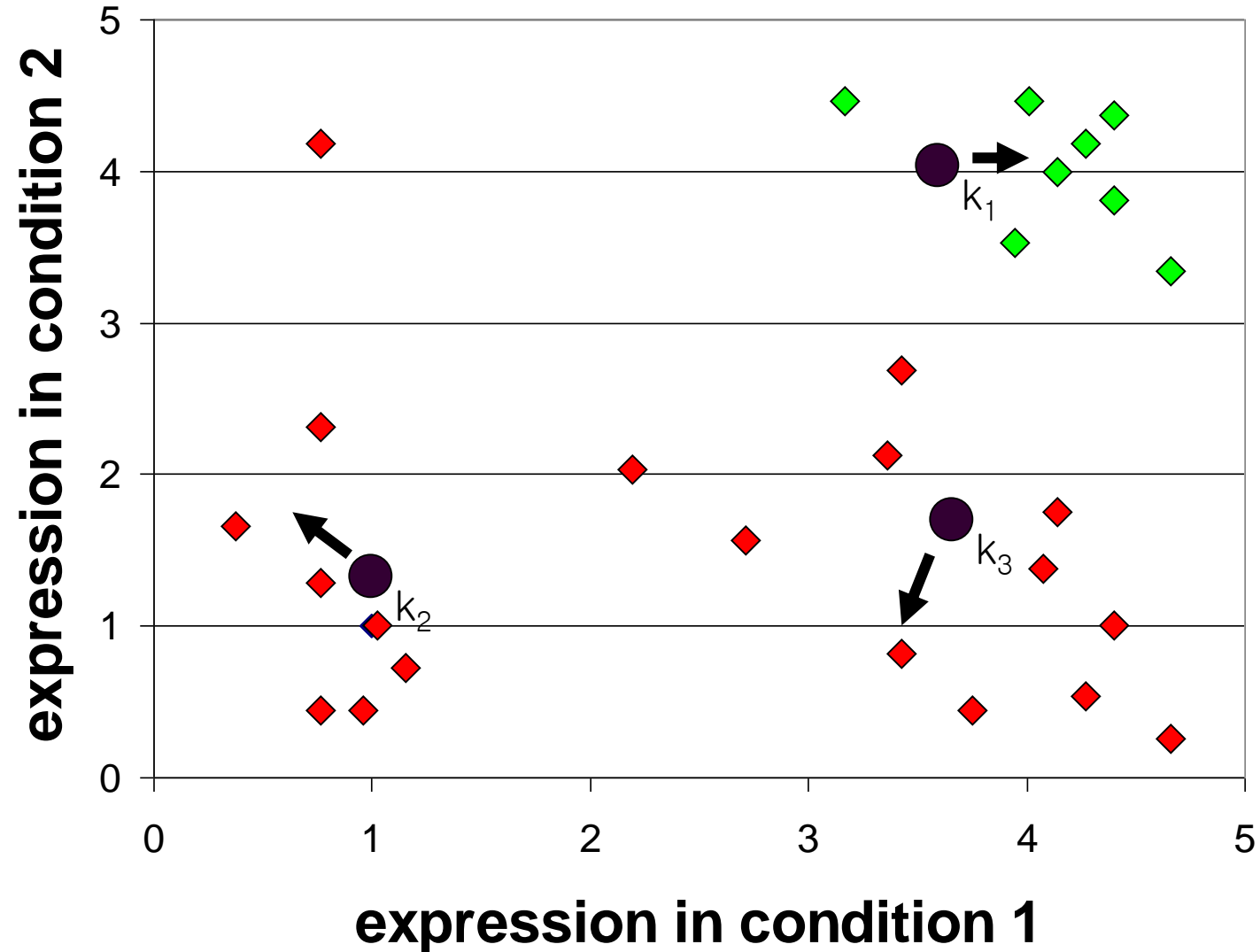
Clustering: Example, Step 3

Algorithm: k-means, Distance Metric: Euclidean Distance



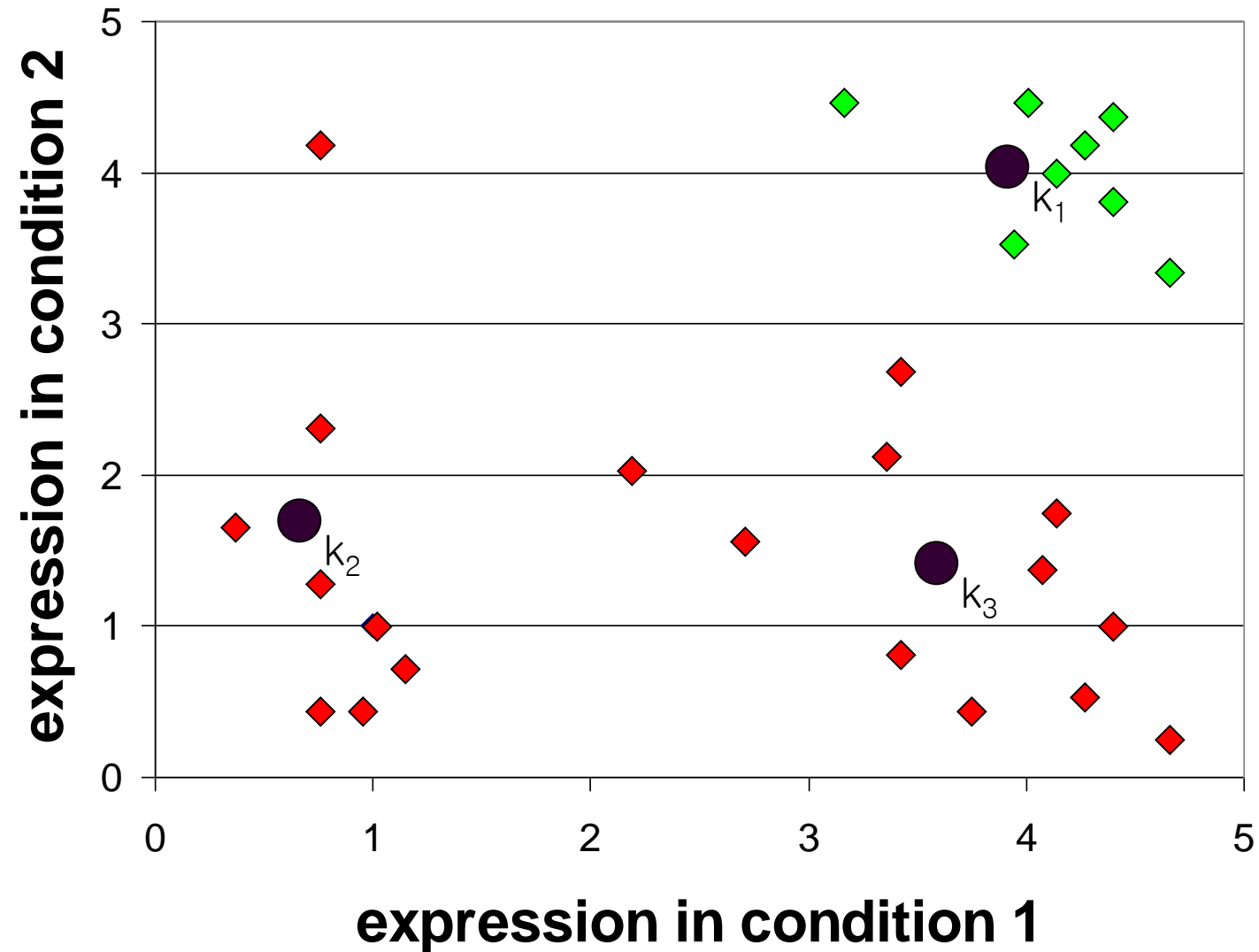
Clustering: Example, Step 4

Algorithm: k-means, Distance Metric: Euclidean Distance



Clustering: Example, Step 5

Algorithm: k-means, Distance Metric: Euclidean Distance



k-평균 알고리즘(k-means Algorithm) 예

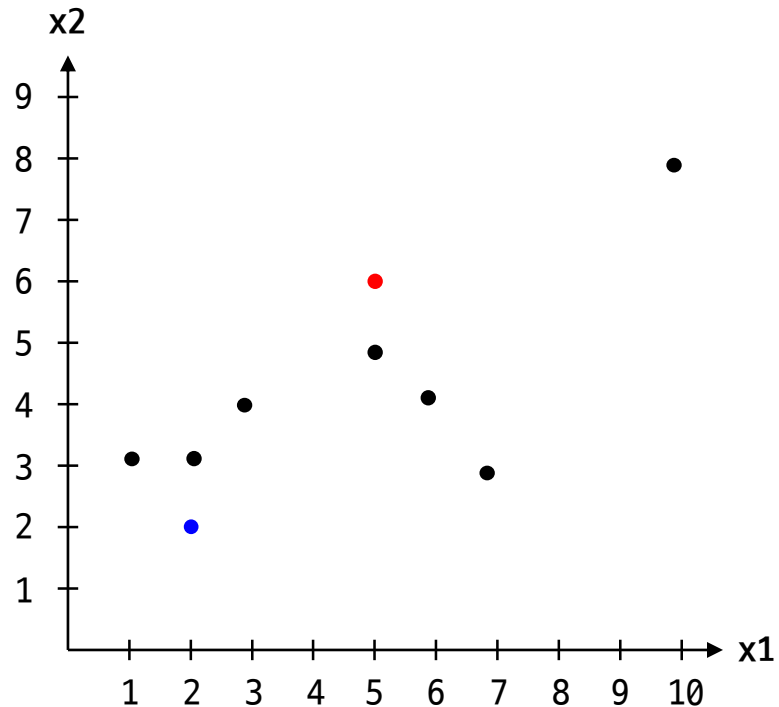
■ Data Set

■ 샘플의 개수 : 9개

- 각 샘플은 2개의 특징(feature)을 가지고 있음

■ Data

- $s_1 = (5, 6)$
- $s_2 = (2, 2)$
- $s_3 = (7, 3)$
- $s_4 = (5, 5)$
- $s_5 = (1, 3)$
- $s_6 = (10, 8)$
- $s_7 = (3, 4)$
- $s_8 = (6, 4)$
- $s_9 = (2, 3)$



초기 cluster center

$$m_1 = (5, 6)$$

$$m_2 = (2, 2)$$

1단계: 초기 센터를 기준으로 분류

■ 각 sample과 cluster center와의 거리 계산

■ $s_1 = (5, 6)$ $s_2 = (2, 2)$ $s_3 = (7, 3)$ $s_4 = (5, 5)$ $s_5 = (1, 3)$

■ $s_6 = (10, 8)$ $s_7 = (3, 4)$ $s_8 = (6, 4)$ $s_9 = (2, 3)$

■ 초기 cluster center

■ $m_1 = (5, 6)$: s_1 s_3 s_4 s_6 s_8

■ $m_2 = (2, 2)$: s_2 s_5 s_7 s_9

$$d(s_1, m_1) = \sqrt{(5-5)^2 + (6-6)^2} = 0.0$$

$$d(s_2, m_1) = \sqrt{(2-5)^2 + (2-6)^2} = 5.0$$

$$d(s_3, m_1) = \sqrt{(7-5)^2 + (3-6)^2} = 3.6$$

$$d(s_4, m_1) = \sqrt{(5-5)^2 + (5-6)^2} = 1.0$$

$$d(s_5, m_1) = \sqrt{(1-5)^2 + (3-6)^2} = 5.0$$

$$d(s_6, m_1) = \sqrt{(10-5)^2 + (8-6)^2} = 5.4$$

$$d(s_7, m_1) = \sqrt{(3-5)^2 + (4-6)^2} = 5.3$$

$$d(s_8, m_1) = \sqrt{(6-5)^2 + (4-6)^2} = 2.2$$

$$d(s_9, m_1) = \sqrt{(2-5)^2 + (3-6)^2} = 4.2$$

$$d(s_1, m_2) = \sqrt{(5-2)^2 + (6-2)^2} = 5.0$$

$$d(s_2, m_2) = \sqrt{(2-2)^2 + (2-2)^2} = 0.0$$

$$d(s_3, m_2) = \sqrt{(7-2)^2 + (3-2)^2} = 5.1$$

$$d(s_4, m_2) = \sqrt{(5-2)^2 + (5-2)^2} = 4.2$$

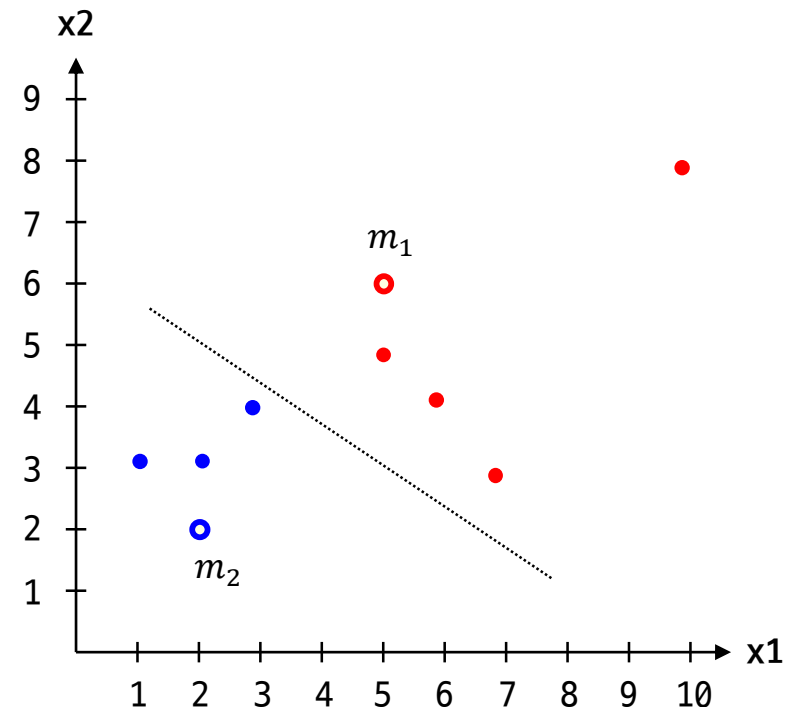
$$d(s_5, m_2) = \sqrt{(1-2)^2 + (3-2)^2} = 1.4$$

$$d(s_6, m_2) = \sqrt{(10-2)^2 + (8-2)^2} = 10.0$$

$$d(s_7, m_2) = \sqrt{(3-2)^2 + (4-2)^2} = 2.2$$

$$d(s_8, m_2) = \sqrt{(6-2)^2 + (4-2)^2} = 4.5$$

$$d(s_9, m_2) = \sqrt{(2-2)^2 + (3-2)^2} = 1.0$$



2단계: 분류결과를 이용하여 센터를 다시 계산

■ 각 그룹 sample들의 mean(평균)을 계산

$$\blacksquare s_1 = (5, 6) \quad s_2 = (2, 2) \quad s_3 = (7, 3) \quad s_4 = (5, 5) \quad s_5 = (1, 3)$$

$$\blacksquare s_6 = (10, 8) \quad s_7 = (3, 4) \quad s_8 = (6, 4) \quad s_9 = (2, 3)$$

■ cluster center 멤버

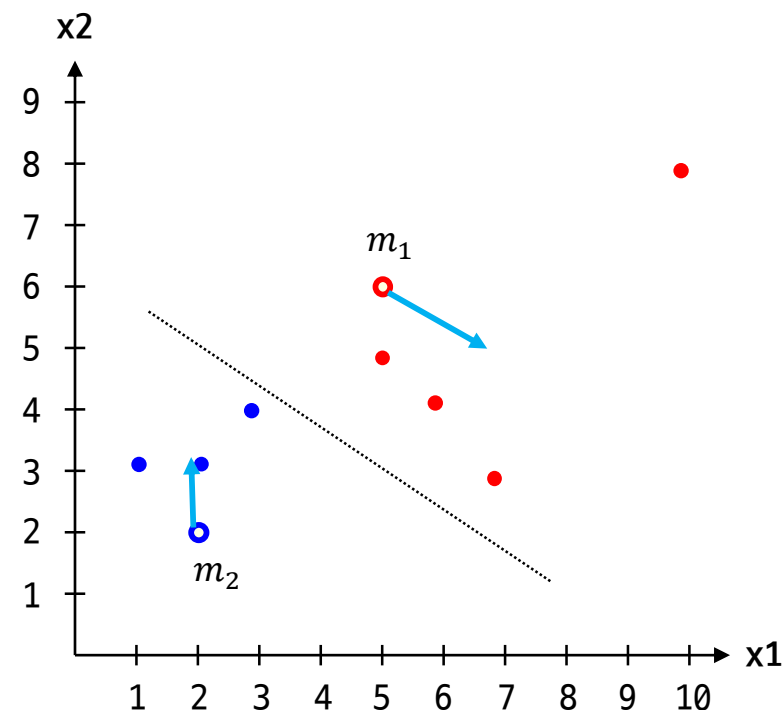
$$\blacksquare m_1: s_1 \ s_3 \ s_4 \ s_6 \ s_8$$

$$\blacksquare m_2: s_2 \ s_5 \ s_7 \ s_9$$

■ cluster center 다시 계산

$$\blacksquare m_1 = \left(\frac{5+7+5+10+6}{5}, \frac{6+3+5+8+4}{5} \right) = (6.6, 5.2)$$

$$\blacksquare m_2 = \left(\frac{2+1+3+2}{4}, \frac{2+3+4+3}{4} \right) = (2.0, 3.0)$$



3단계: 단계 1, 2를 반복

- cluster center의 변화가 ϵ 보다 작을 때까지 단계 1~2를 반복

- Data

- $s_1 = (5, 6)$ $s_2 = (2, 2)$ $s_3 = (7, 3)$ $s_4 = (5, 5)$ $s_5 = (1, 3)$

- $s_6 = (10, 8)$ $s_7 = (3, 4)$ $s_8 = (6, 4)$ $s_9 = (2, 3)$

- cluster center 멤버

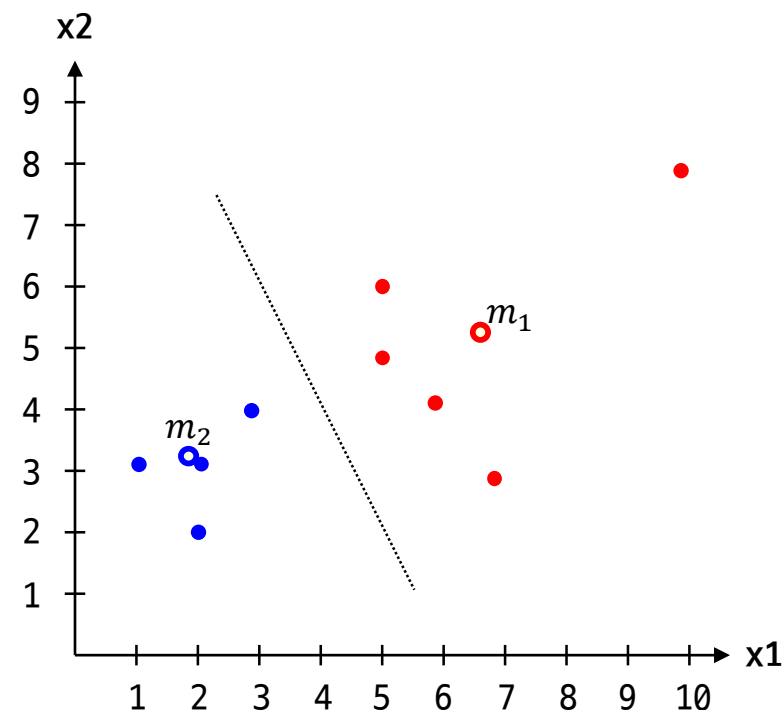
- $m_1: s_1 s_3 s_4 s_6 s_8$

- $m_2: s_2 s_5 s_7 s_9$

- 최종 cluster center

- $m_1 = \left(\frac{5+7+5+10+6}{5}, \frac{6+3+5+8+4}{5} \right) = (6.6, 5.2)$

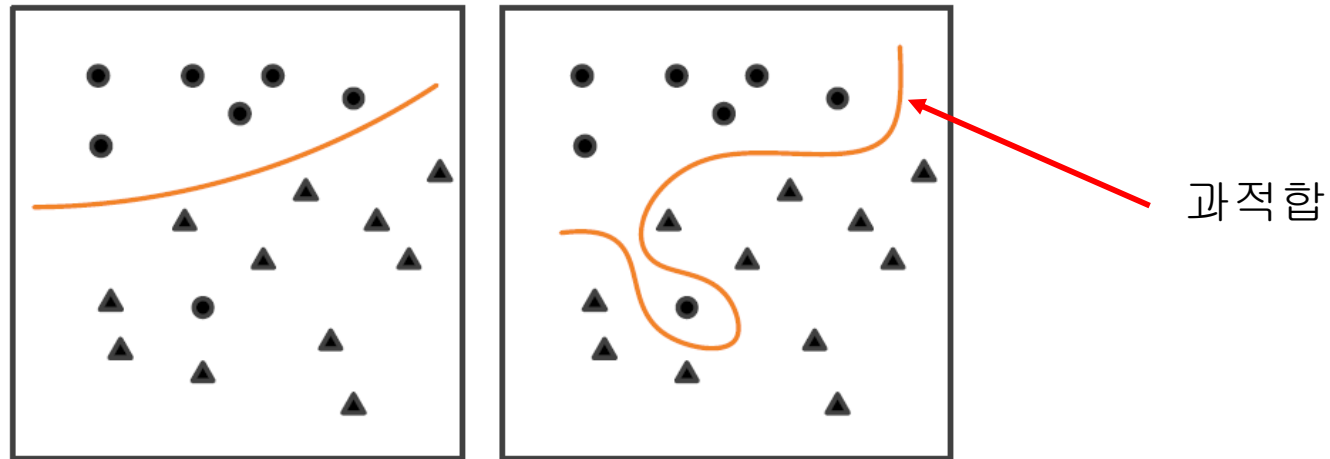
- $m_2 = \left(\frac{2+1+3+2}{4}, \frac{2+3+4+3}{4} \right) = (2.0, 3.0)$



참고 :

■ 학습의 일반화능력

- 학습 과정에서 사용하지 않은 테스트집합으로 학습이 완료된 분류기 성능을 평가
- 학습집합과 테스트집합에 대한 성능이 비슷하면 일반화 능력이 뛰어나
- 과적합 하면 일반화 능력이 떨어짐



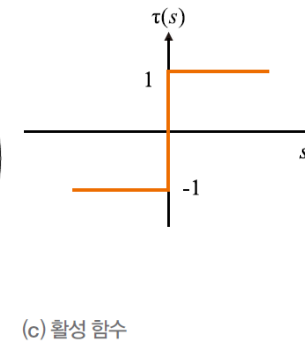
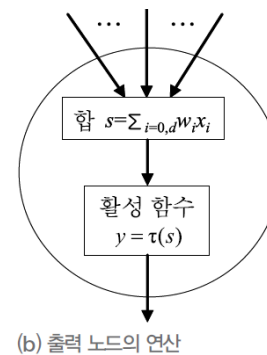
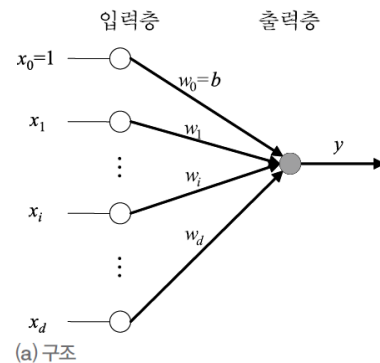
신경망(Neural Network)

■ 뇌 구조를 모방한 계산 모형

- 1950년대 Rosenblatt의 퍼셉트론
- 1980년대 퍼셉트론을 확장한 다층퍼셉트론(MLP)
- 일반화 능력이 뛰어남

■ 퍼셉트론(Perceptron)

- 입력은 특징 벡터 $\mathbf{x}=(x_1, x_2, \dots, x_d)$
- \mathbf{x} 를 두 개의 부류 ω_1 과 ω_2 중의 하나로 분류하는 이진 분류기



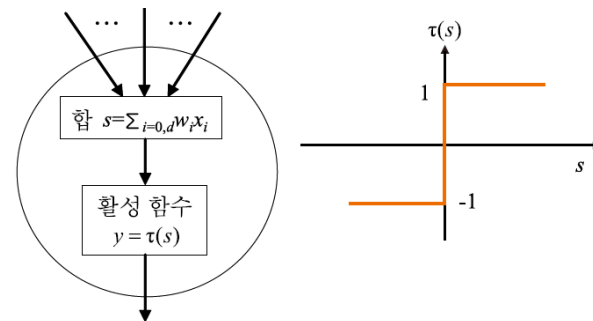
퍼셉트론

■ 출력 노드는 가중치 합과 활성화 함수 계산

- 특징 벡터 $x=(x_1, x_2, \dots, x_d)$, 가중치 벡터 $w=(w_1, w_2, \dots, w_d)$ 로 표기하면

$$y = \tau(s) = \tau\left(\sum_{i=1}^d w_i x_i + b\right) = \tau(\mathbf{w}\mathbf{x}^T + b)$$

$$\text{이때 } \tau(s) = \begin{cases} +1, & s \geq 0 \\ -1, & s < 0 \end{cases}$$



(b) 출력 노드의 연산

(c) 활성화 함수

■ 계단 함수를 활성화 함수로 사용

- 출력은 +1(w_1 에 해당) 또는 -1(w_2 에 해당)
- 이진 분류기

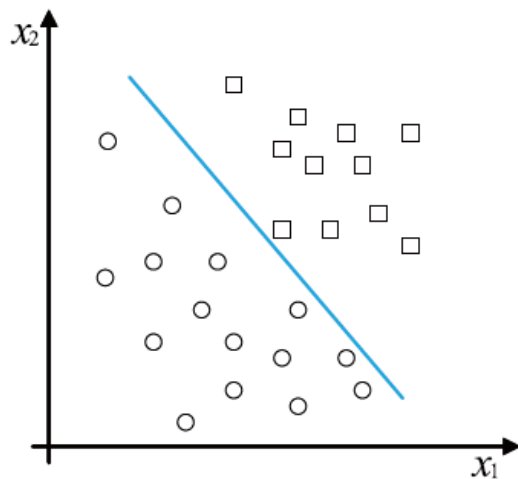
퍼셉트론의 한계

■ 한계

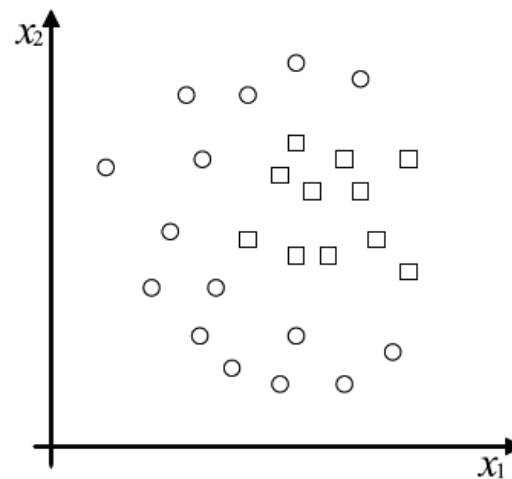
- 퍼셉트론은 결정초평면 역할을 하는 선형분류기
- 선형 분리 불가능한 상황에 대처하지 못함

$$d(\mathbf{x}) = \mathbf{w}\mathbf{x}^T + b \geq 0 \text{이면 } \mathbf{x} \in \omega_1$$

$$d(\mathbf{x}) = \mathbf{w}\mathbf{x}^T + b < 0 \text{이면 } \mathbf{x} \in \omega_2$$



(a) 선형 분리 가능

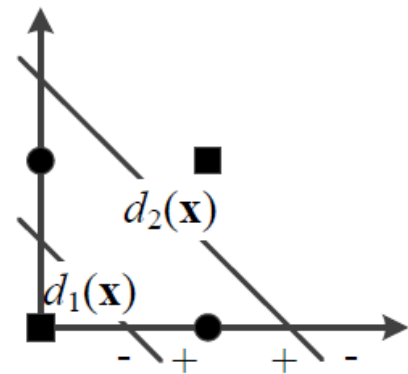


(b) 선형 분리 불가능

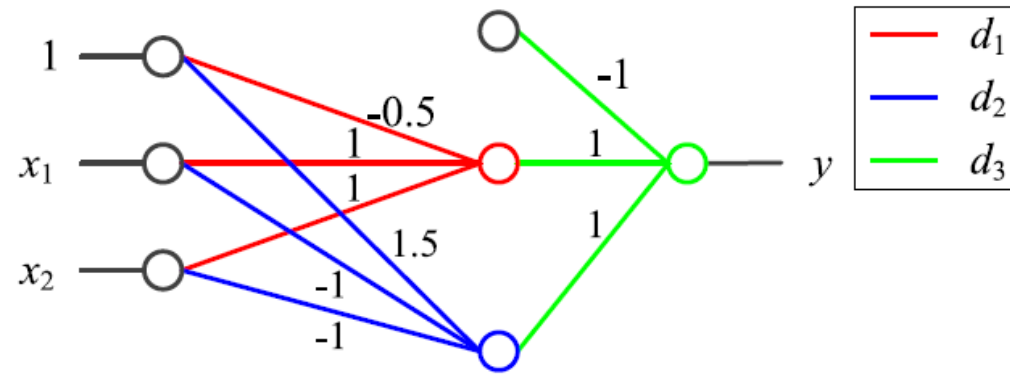
다층 퍼셉트론(MLP)으로 확장

■ XOR 분류 문제

- 단일 Perceptron으로는 분류 불가능(75%한계)
- 3개의 Perceptron(NAND, OR, AND)으로 해결
 - d_1 의 +영역과 d_2 의 +영역이 겹친 영역은 ω_1 , 나머지는 ω_2



(a) XOR 분류 문제

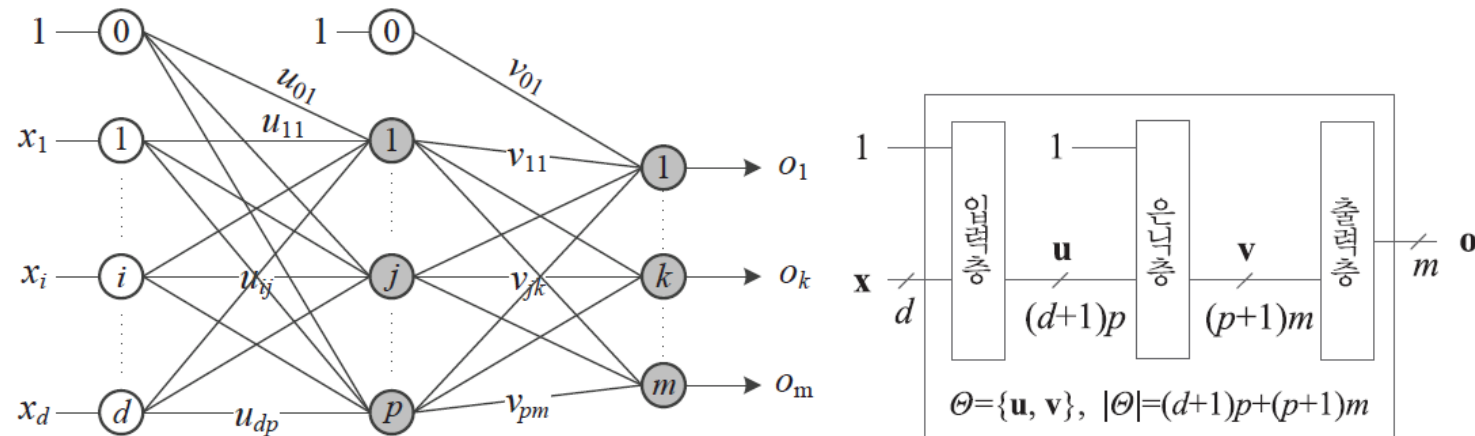


(b) 세 개 퍼셉트론의 결합

■ 다층 퍼셉트론의 구조

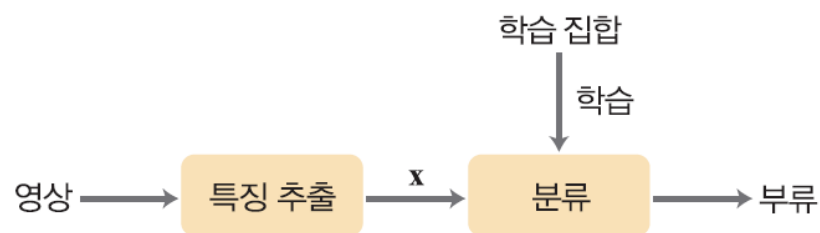
■ 입력층 → 은닉층 → 출력층

- 입력층: 특징 벡터의 차원에 따라 d 개의 노드와 여분의 바이어스 노드로 구성
- 출력층: 부류 개수에 따라 m 개의 노드로 구성
- 은닉층: 노드 개수 p 를 사용자가 설정

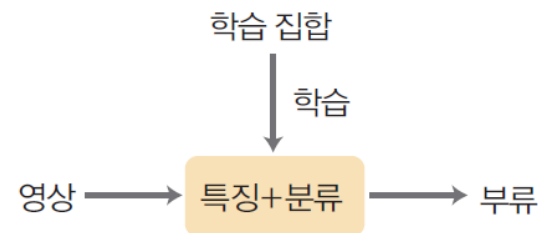


깊은 학습(Deep Learning)

- MLP의 은닉층 개수를 늘린 깊은 신경망
 - 층이 깊어 MLP가 사용하는 오류 역전파 알고리즘만으로 학습 불가능
 - 1986년 Hinton에 의해 해결됨
- CNN(Convolutional Neural Network)
 - 과거 : 특징추출을 따로 설계하고 구현
 - CNN : 특징추출과 분류를 하나의 학습모델로 처리



(a) 사람이 설계한 특징+학습에 의한 분류



(b) 학습에 의한 특징+분류

■ 컨볼루션 신경망 - LeCun 1998

- 보통 특징은 고정된 마스크로 컨볼루션을 수행함으로써 추출 (예) 에지 마스크

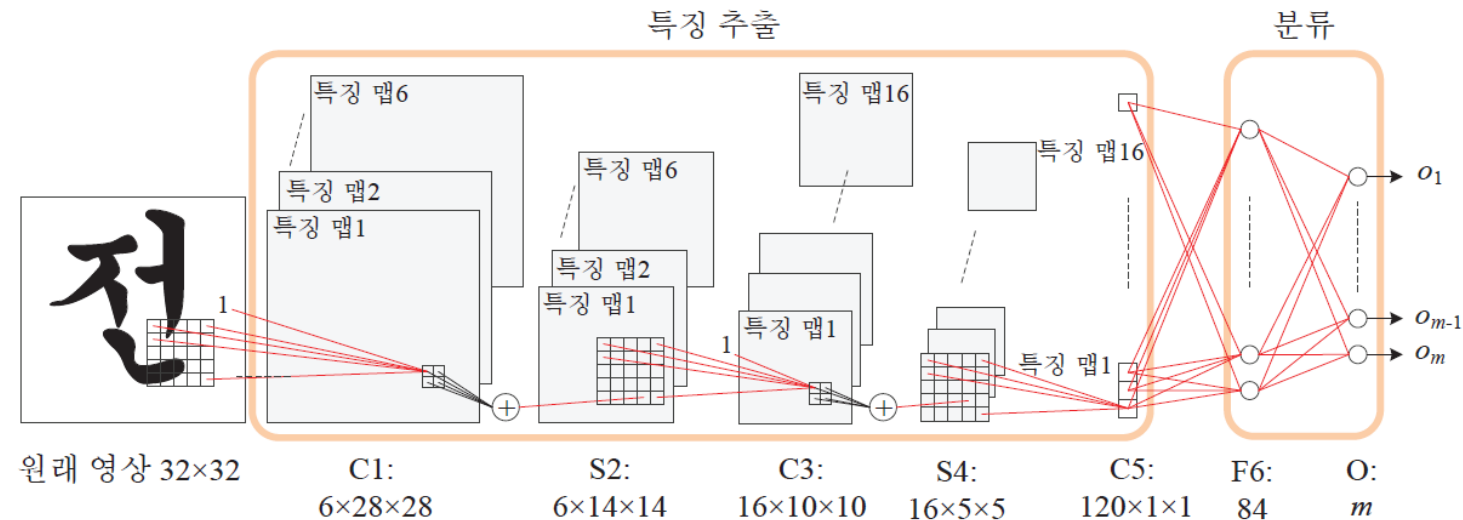
<table><tr><td>0</td><td>-1</td></tr><tr><td>1</td><td>0</td></tr></table> m_y	0	-1	1	0	<table><tr><td>-1</td><td>0</td></tr><tr><td>0</td><td>1</td></tr></table> m_x	-1	0	0	1	<table><tr><td>-1</td><td>-1</td><td>-1</td></tr><tr><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>1</td><td>1</td></tr></table> m_y	-1	-1	-1	0	0	0	1	1	1	<table><tr><td>-1</td><td>0</td><td>1</td></tr><tr><td>-1</td><td>0</td><td>1</td></tr><tr><td>-1</td><td>0</td><td>1</td></tr></table> m_x	-1	0	1	-1	0	1	-1	0	1	<table><tr><td>-1</td><td>-2</td><td>-1</td></tr><tr><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>2</td><td>1</td></tr></table> m_y	-1	-2	-1	0	0	0	1	2	1	<table><tr><td>-1</td><td>0</td><td>1</td></tr><tr><td>-2</td><td>0</td><td>2</td></tr><tr><td>-1</td><td>0</td><td>1</td></tr></table> m_x	-1	0	1	-2	0	2	-1	0	1
0	-1																																																
1	0																																																
-1	0																																																
0	1																																																
-1	-1	-1																																															
0	0	0																																															
1	1	1																																															
-1	0	1																																															
-1	0	1																																															
-1	0	1																																															
-1	-2	-1																																															
0	0	0																																															
1	2	1																																															
-1	0	1																																															
-2	0	2																																															
-1	0	1																																															
(a) 로버츠		(b) 프레윗		(c) 소벨																																													

- 주어진 문제에 가장 적합한 마스크를 학습으로 알아내자! ← 컨볼루션 신경망의 발상!

■ 컨볼루션 신경망의 구조

■ 일곱 층으로 구성

- 앞 단 C1-S2-C3-S4-C5는 특징 추출 담당 (C층은 컨볼루션, S는 다운샘플링 수행)
- 뒤 단 F6-O는 분류 담당(MLP와 같은 구조)



■ 깊은 신경망의 우수성

- MNIST 데이터베이스에서 우승
(<http://yann.lecun.com/exdb/mnist/>)

3 6 8 1 7 9 6 6 9 1
6 7 5 7 8 6 3 4 8 5
2 1 7 9 7 1 2 8 4 5

- 1000부류의 자연 영상을 인식하는 ILSVRC 대회에서 우승

