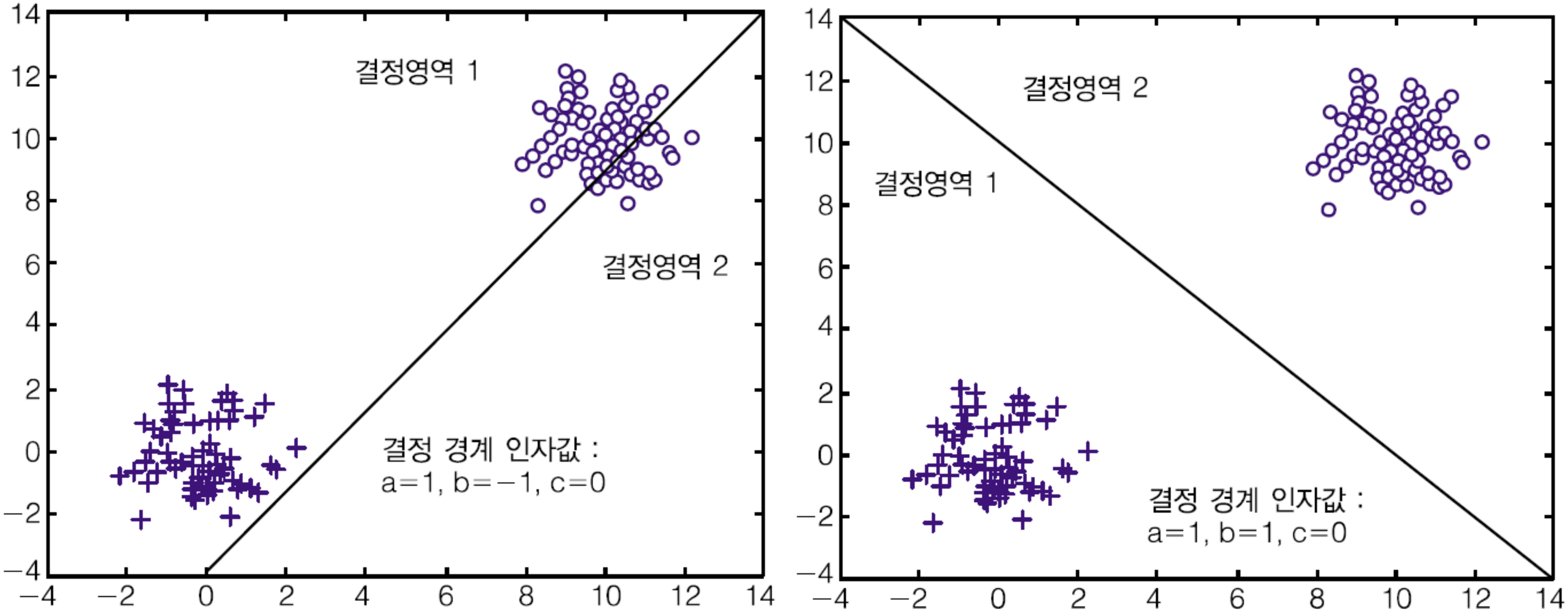
01_선형 분류기

❖ 선형 분류기(linear classifier)

$$g(\mathbf{x}, \mathbf{y}) = a\mathbf{x} + b\mathbf{y} + c$$

※ a, b, c는 가중치(weights)

가중치의 개수는(특징 벡터의 차원 수+1)



(a) 학습 전의 결정 경계 (b) 학습 후의 결정 경계

01_선형 분류기

❖ 판별식 가중치 결정

- 판별식을 정의하는데 중요한 파라미터 : 가중치벡터 w

- 라벨이 정해진 N 차원의 특징벡터집합인 학습 데이터로부터 최소의 분류 오차를 생성하는 최적의 가중치 파라미터를 찾음

$$g(\mathbf{x}, \cancel{\mathbf{y}}) = < \mathbf{w}, \mathbf{x} > + \mathbf{w}_0$$

$$= \sum_{i=1}^{n} \mathbf{w}_i \mathbf{x}_i + \mathbf{w}_0$$

- $\mathbf{w} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n)^T$ 는 가중치 벡터를 의미한다

- $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)^T$ 는 입력 벡터다.

01_선형 분류기

❖ 판별식 가중치 결정

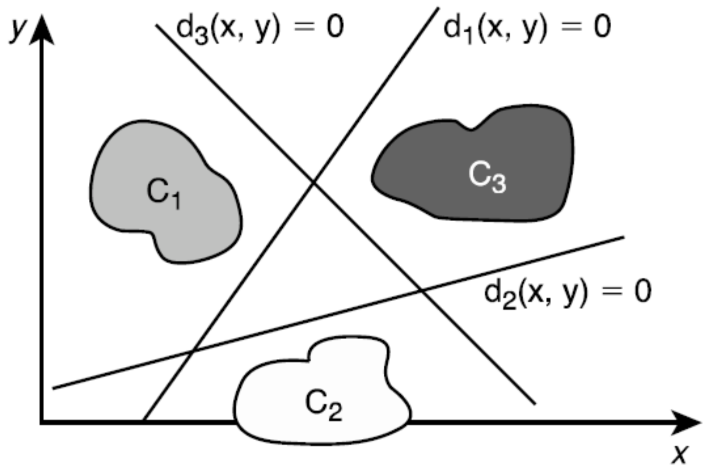
- 가중치 벡터 w는 판별식을 결정하는 데 중요한 파라미터다. 그렇다면 이 파라미터를 어떻게 학습할 것인가?
- 학습 데이터가 클래스 라벨이 붙은 N차원의 특징 벡터 집합일 때, 분류오차를 최소로 하는 **최적의 가중치 파라미터를 찾아야** 한다.
- 이 가중치를 찾는 여러 가지 접근법이 있는데, 대부분 주어진 학습 집합에서 **분류 오차를 최소화**하는 방식을 채택하고 있다.

❖ 결정 경계

- 두 개의 클래스에 대한 문제에서는 결정 경계(decision boundary)로 선형 판별식 한 개가 필요하다.
- 그럼 3개 이상의 클래스를 분류하려면, 몇 개의 선형 판별식이 필요한가?

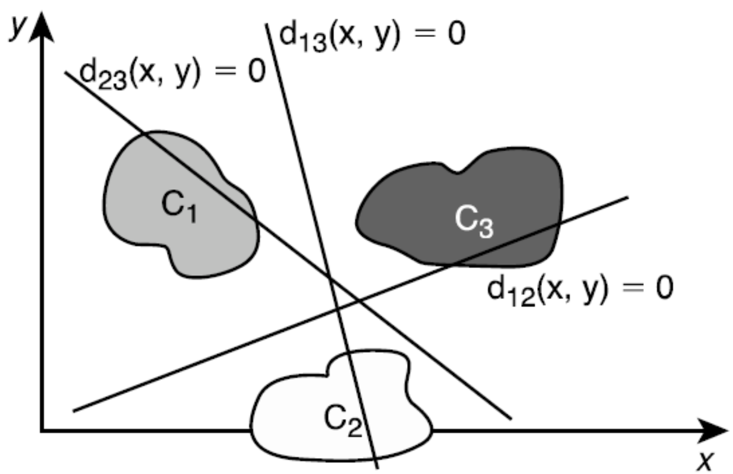
❖ 다중 클래스 결정 경계

- 각 클래스가 단일한 판별식에 의해 결정되는 경우



[그림 6-2] 단일한 판별식

- 판별식이 클래스의 쌍으로 결정되는 경우



[그림 6-3] 쌍으로 분류되는 판별식

- 3개의 클래스와 3개의 판별식이 주어진 경우

만약 $d_1(x, y) > d_2(x, y)$고 $d_1(x, y) > d_3(x, y)$면, 클래스 1로 결정한다.

만약 $d_2(x, y) > d_1(x, y)$고 $d_2(x, y) > d_3(x, y)$면, 클래스 2로 결정한다.

만약 $d_3(x, y) > d_2(x, y)$고 $d_3(x, y) > d_1(x, y)$면, 클래스 3으로 결정한다.

02_이차 분류기

- 이차 분류기: 판별함수가 가우시안 분포를 이룬다고 가정할 경우
- 선형 분류기: 편별식은 2차원에서는 직선형태, 3차원에서는 평면형태
- 최소-거리 분류기 혹은 마할나노비스(mahalanobis) 분류기 조건
  - 모두 가우시안 분포에 따르며
  - 같은 공분산 값을 가지며
  - 사전 확률이 같을 경우 사용하는 방법
- 유클리디안 거리 분류기 조건
  - 모두 가우시안 분포에 따르며
  - 항등 행렬에 비례하는 동일한 공분산 값을 가지며
  - 사전 확률이 같을 경우

$$\text{Case } 1: \ \mathbf{\Sigma_i} = \boldsymbol{\sigma}^2\mathbf{I}$$

$$\text{Case } 2: \ \mathbf{\Sigma_i} = \mathbf{\Sigma} \ \ (\Sigma: \text{ 대각행렬})$$

$$\text{Case } 3: \ \mathbf{\Sigma_i} = \mathbf{\Sigma} \ \ (\Sigma: \text{ 비대각행렬})$$

$$\text{Case } 4: \ \mathbf{\Sigma_i} = \boldsymbol{\sigma}_\mathbf{i}^2\mathbf{I}$$

$$\text{Case } 5: \ \mathbf{\Sigma_i} \neq \mathbf{\Sigma_j} \ \ (\text{일반형})$$

❖ 가우시안 확률밀도함수의 일반식

$$f_x(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right]$$

$$g_i(x) = P(\omega_i \mid x) = \frac{P(x \mid \omega_i)P(\omega_i)}{P(x)}$$

$$= \frac{1}{(2\pi)^{n/2}|\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\right] P(\omega_i)\frac{1}{P(x)}$$

$$g_i(x) = |\Sigma_i|^{-1/2} \exp\left[-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\right] P(\omega_i)$$

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) - \frac{1}{2}\log(|\Sigma_i|) + \log(P(\omega_i))$$

❖ 데이터에 대한 통계적인 공분산 행렬의 종류

Case 1: $\Sigma_i = \sigma^2 I$

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T(\sigma^2 I)^{-1}(x-\mu_i) - \frac{1}{2}\log(|\sigma^2 I|) + \log(P(\omega_i))$$

$$= -\frac{1}{2\sigma^2}(x-\mu_i)^T(x-\mu_i) - \frac{1}{2}N\log(\sigma^2) + \log(P(\omega_i))$$

$$= -\frac{1}{2\sigma^2}(x-\mu_i)^T(x-\mu_i) + \log(P(\omega_i))$$

$$g_i(x) = -\frac{1}{2\sigma^2}(x-\mu_i)^T(x-\mu_i) + \log(P(\omega_i))$$

$$= -\frac{1}{2\sigma^2}(x^T x - 2\mu_i^T x + \mu_i^T \mu_i) + \log(P(\omega_i))$$

❖ 데이터에 대한 통계적인 공분산 행렬의 종류

Case 1: $\Sigma_i = \sigma^2 \mathbf{I}$

$$g_i(x) = -\frac{1}{2\sigma^2}(-2\mu_i^T x + \mu_i^T \mu_i) + \log(P(\omega_i)) = w_i^T x + w_{i0}$$

$$※ \begin{cases} w_i = \dfrac{\mu_i}{\sigma^2} \\ w_{i0} = -\dfrac{1}{2\sigma^2}\mu_i^T \mu_i + \log(P(\omega_i)) \end{cases}$$

$$g_i(x) = -\frac{1}{2\sigma^2}(x - \mu_i)^T (x - \mu_i)$$

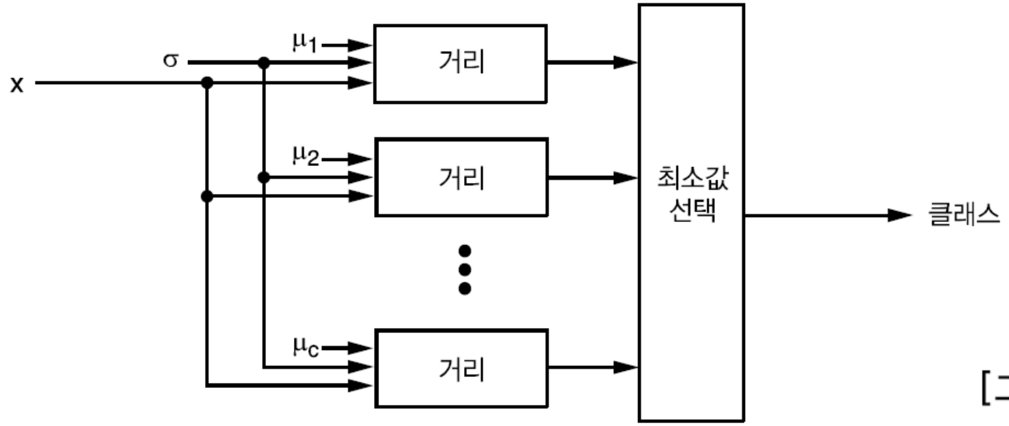
❖ 데이터에 대한 통계적인 공분산 행렬의 종류

예 다음과 같은 클래스 평균과 공분산 행렬에 따르며, 2차원 특징 벡터로 이루어진 3개의 클래스에 대한 결정 경계를 계산한다(사전 확률은 동일하다고 가정).

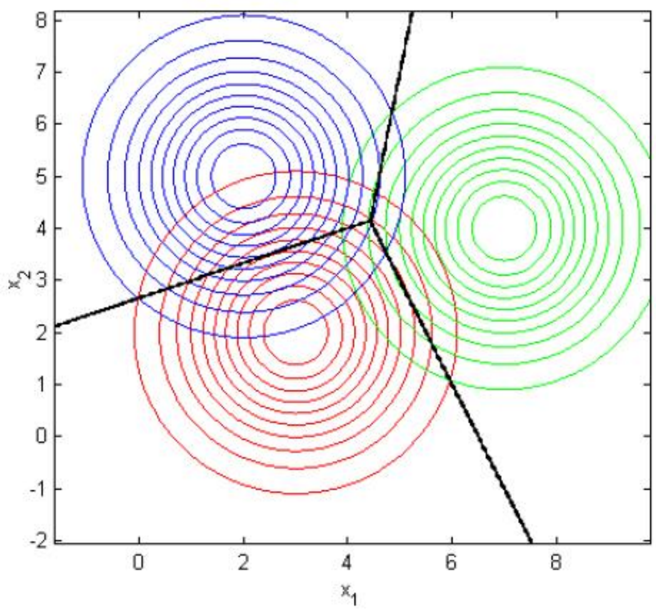
$$\mu_1 = \begin{bmatrix} 3 & 2 \end{bmatrix}^T \qquad \mu_2 = \begin{bmatrix} 7 & 4 \end{bmatrix}^T \qquad \mu_3 = \begin{bmatrix} 2 & 5 \end{bmatrix}^T$$

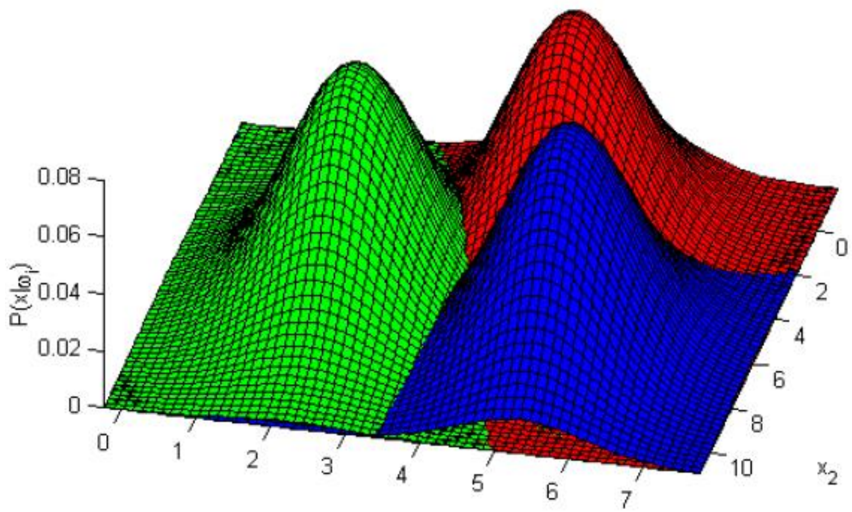
$$\Sigma_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \qquad \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

❖ 데이터에 대한 통계적인 공분산 행렬의 종류



[그림 6-4] 최근접 평균 분류기





[그림 6-5] Case 1의 특징 벡터의 분포와 결정 경계

❖ 데이터에 대한 통계적인 공분산 행렬의 종류

Case 2: $\Sigma_i = \Sigma$ (Σ : 대각행렬)

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) - \frac{1}{2}\log(|\Sigma_i|) + \log(P(\omega_i))$$

$$= -\frac{1}{2}(x-\mu_i)^T \begin{bmatrix} \sigma_1^{-2} & & \\ & \ddots & \\ & & \sigma_N^{-2} \end{bmatrix}^{-1} (x-\mu_i) - \frac{1}{2}\log\left(\begin{vmatrix} \sigma_1^{-2} & & \\ & \ddots & \\ & & \sigma_N^{-2} \end{vmatrix}\right) + \log(P(\omega_i))$$

$$= -\frac{1}{2}\sum_{k=1}^{N}\frac{(x[k]-\mu_i[k])^2}{\sigma_k^2} - \frac{1}{2}\log\prod_{k=1}^{N}\sigma_k^2 + \log(P(\omega_i))$$

$$= -\frac{1}{2}\sum_{k=1}^{N}\frac{x[k]^2 - 2x[k]\mu_i[k] + \mu_i[k]^2}{\sigma_k^2} - \frac{1}{2}\log\prod_{k=1}^{N}\sigma_k^2 + \log(P(\omega_i))$$

❖ 데이터에 대한 통계적인 공분산 행렬의 종류

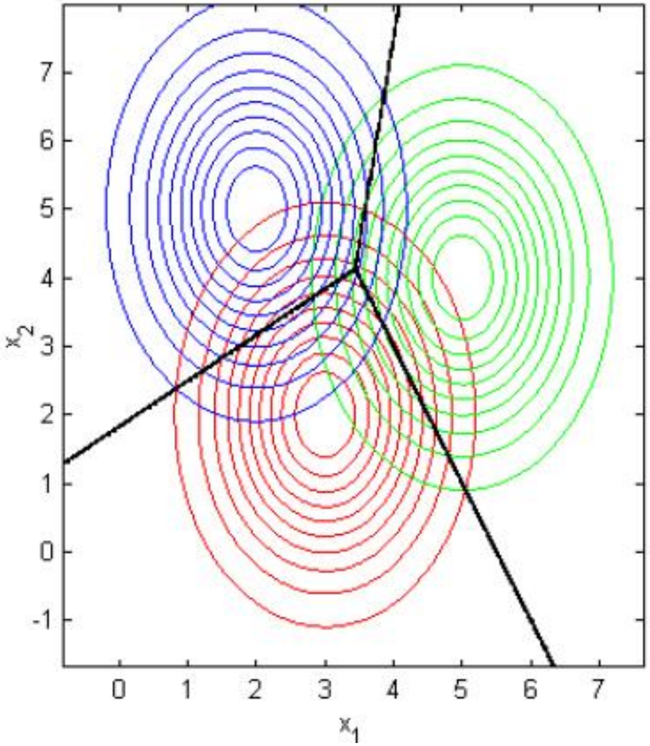
예 다음과 같은 클래스 평균과 공분산 행렬에 따르면, 2차원 특징 벡터로 이루어진 3개의 클래스에 대한 결정 경계를 계산한다(사전 확률은 동일하다고 가정).

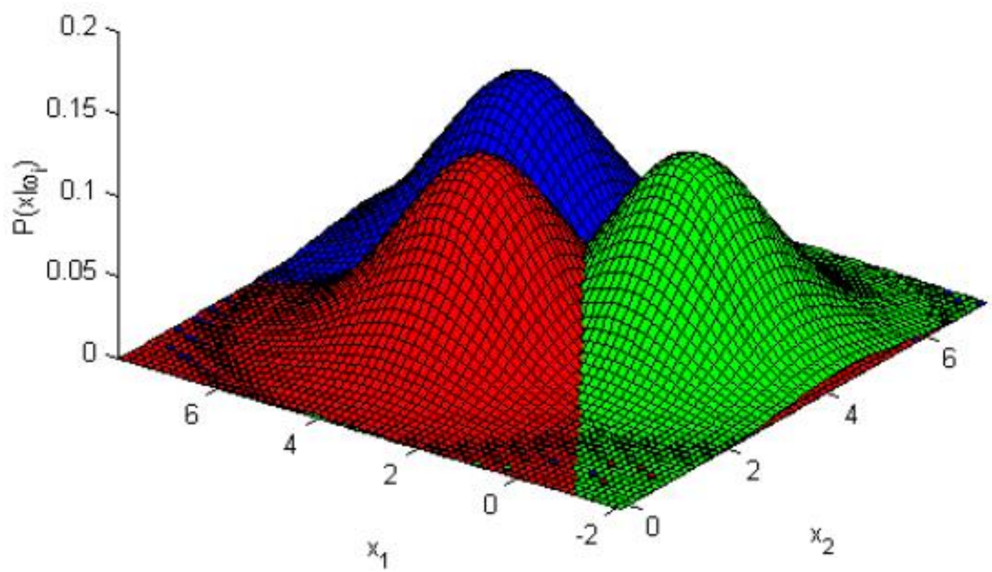
$$\mu_1 = \begin{bmatrix} 3 & 2 \end{bmatrix}^T \qquad \mu_2 = \begin{bmatrix} 5 & 4 \end{bmatrix}^T \qquad \mu_3 = \begin{bmatrix} 2 & 5 \end{bmatrix}^T$$

$$\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \qquad \Sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

❖ 데이터에 대한 통계적인 공분산 행렬의 종류

$$g_i(x) = -\frac{1}{2}\sum_{k=1}^{N}\frac{-2x[k]\mu_i[k]+\mu_i[k]^2}{\sigma_k^2} - \frac{1}{2}\log\prod_{k=1}^{N}\sigma_k^2 + \log(P(\omega_i))$$





[그림 6-6] Case 2의 특징 벡터의 분포와 결정 경계

❖ 데이터에 대한 통계적인 공분산 행렬의 종류

Case 3: $\Sigma_i = \Sigma$ (Σ: 비대각행렬)

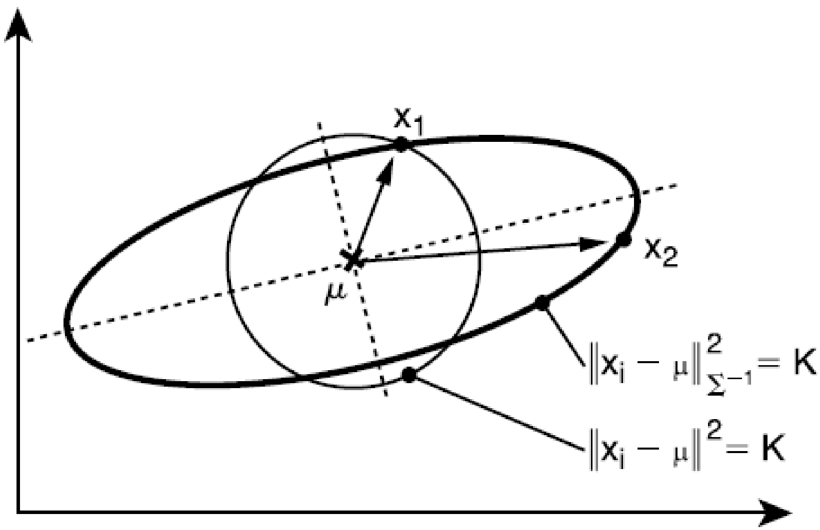
$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) - \frac{1}{2}\log(|\Sigma_i|) + \log(P(\omega_i))$$

$$= -\frac{1}{2}(x-\mu_i)^T \Sigma^{-1}(x-\mu_i) - \frac{1}{2}\log(|\Sigma|) + \log(P(\omega_i))$$

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma^{-1}(x-\mu_i) + \log(P(\omega_i))$$

❖ 마할라노비스 거리

$$\|\mathbf{x}-\mathbf{y}\|^2_{\Sigma^{-1}} = (\mathbf{x}-\mathbf{y})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{y})$$



[그림 6-7] 마할라노비스 거리

❖ 마할라노비스 거리

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i) + \log(P(\omega_i))$$

$$= -\frac{1}{2}(x^T\Sigma^{-1}x - 2\mu_i^T\Sigma^{-1}x + \mu_i^T\Sigma^{-1}\mu_i) + \log(P(\omega_i))$$

$$g_i(x) = -\frac{1}{2}(-2\mu_i^T\Sigma^{-1}x + \mu_i^T\Sigma^{-1}\mu_i) + \log(P(\omega_i))$$

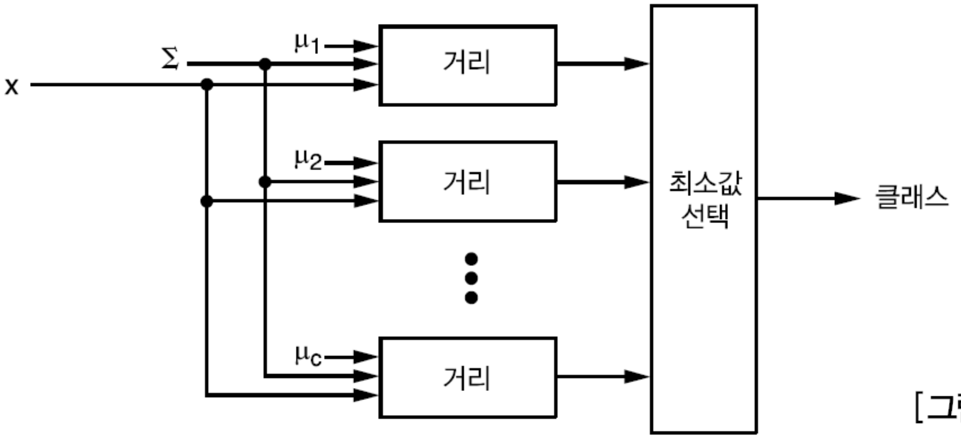
$$g_i(x) = w_i^T x + w_{i0}$$

※ $$\begin{cases} w_i = \Sigma^{-1}\mu_i \\ w_{i0} = -\frac{1}{2}\mu_i^T\Sigma^{-1}\mu_i + \log(P(\omega_i)) \end{cases}$$

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i)$$

02_이차 분류기

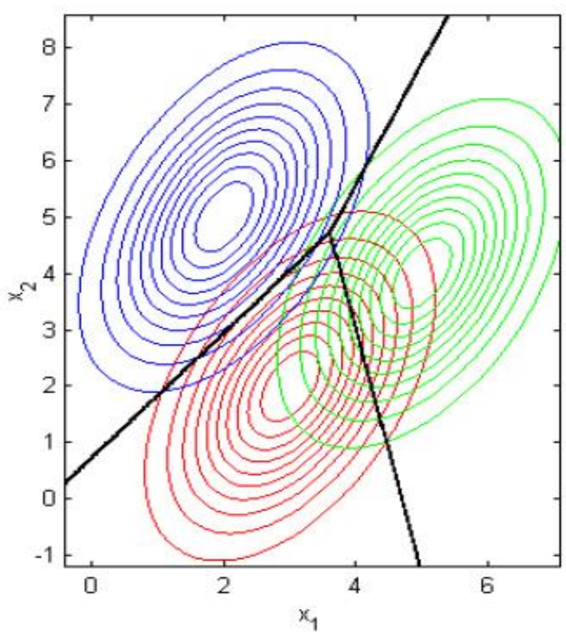
❖ 마할라노비스 거리

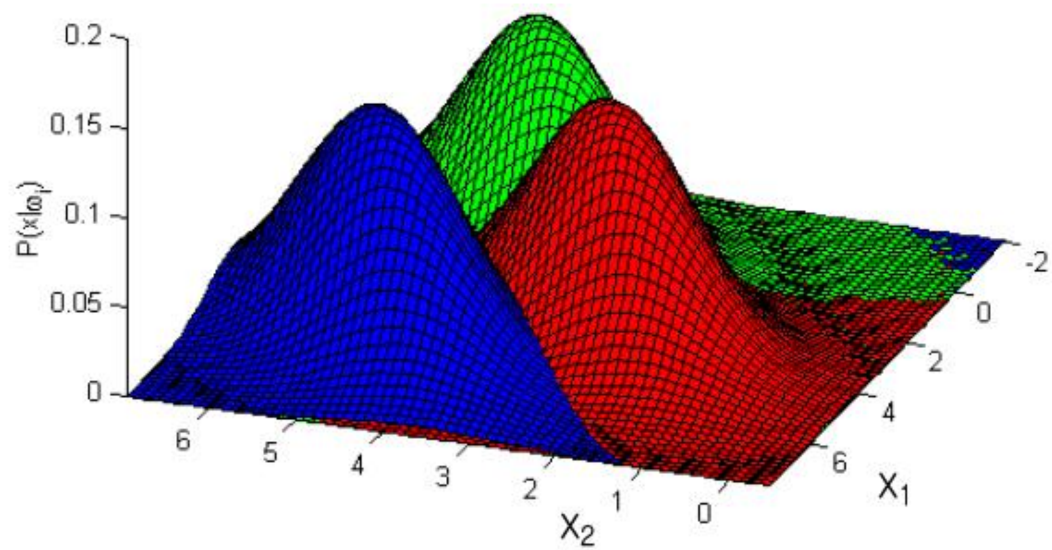


$$\mu_1 = [3 \quad 2]^T \qquad \mu_2 = [5 \quad 4]^T \qquad \mu_3 = [2 \quad 5]^T$$

$$\Sigma_1 = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 2 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 2 \end{bmatrix} \quad \Sigma_3 = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 2 \end{bmatrix}$$

[그림 6-8] 마할라노비스 분류기





[그림 6-9] Case 3의 특징 벡터의 분포와 결정 경계

❖ 데이터에 대한 통계적인 공분산 행렬의 종류

Case 4: $\Sigma_i = \sigma_i^2 \mathrm{I}$

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) - \frac{1}{2}\log(|\Sigma_i|) + \log(P(\omega_i))$$

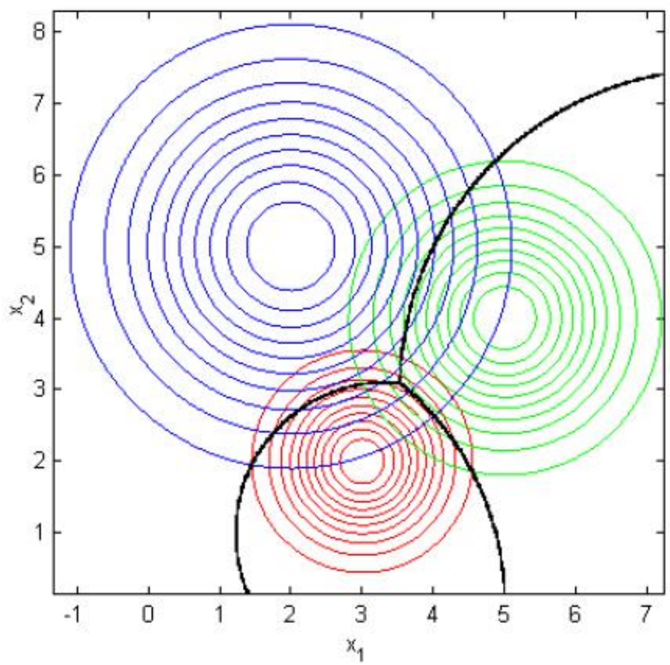
$$= -\frac{1}{2}(x-\mu_i)^T \sigma_i^{-2}(x-\mu_i) - \frac{1}{2} N\log(\sigma_i^2) + \log(P(\omega_i))$$

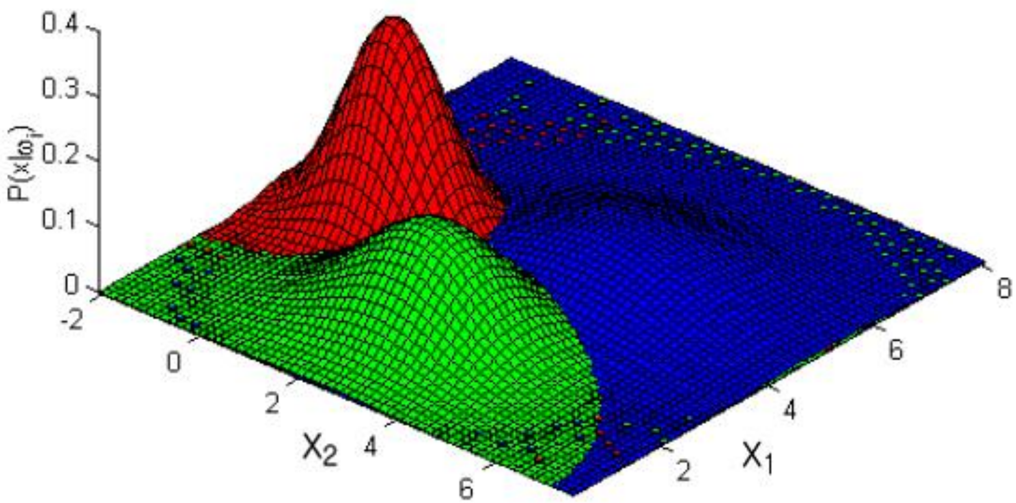
❖ 데이터에 대한 통계적인 공분산 행렬의 종류

Case 4: $\Sigma_i = \sigma_i^2 I$

$$\mu_1 = [3 \quad 2]^T \qquad \mu_2 = [5 \quad 4]^T \qquad \mu_3 = [2 \quad 5]^T$$

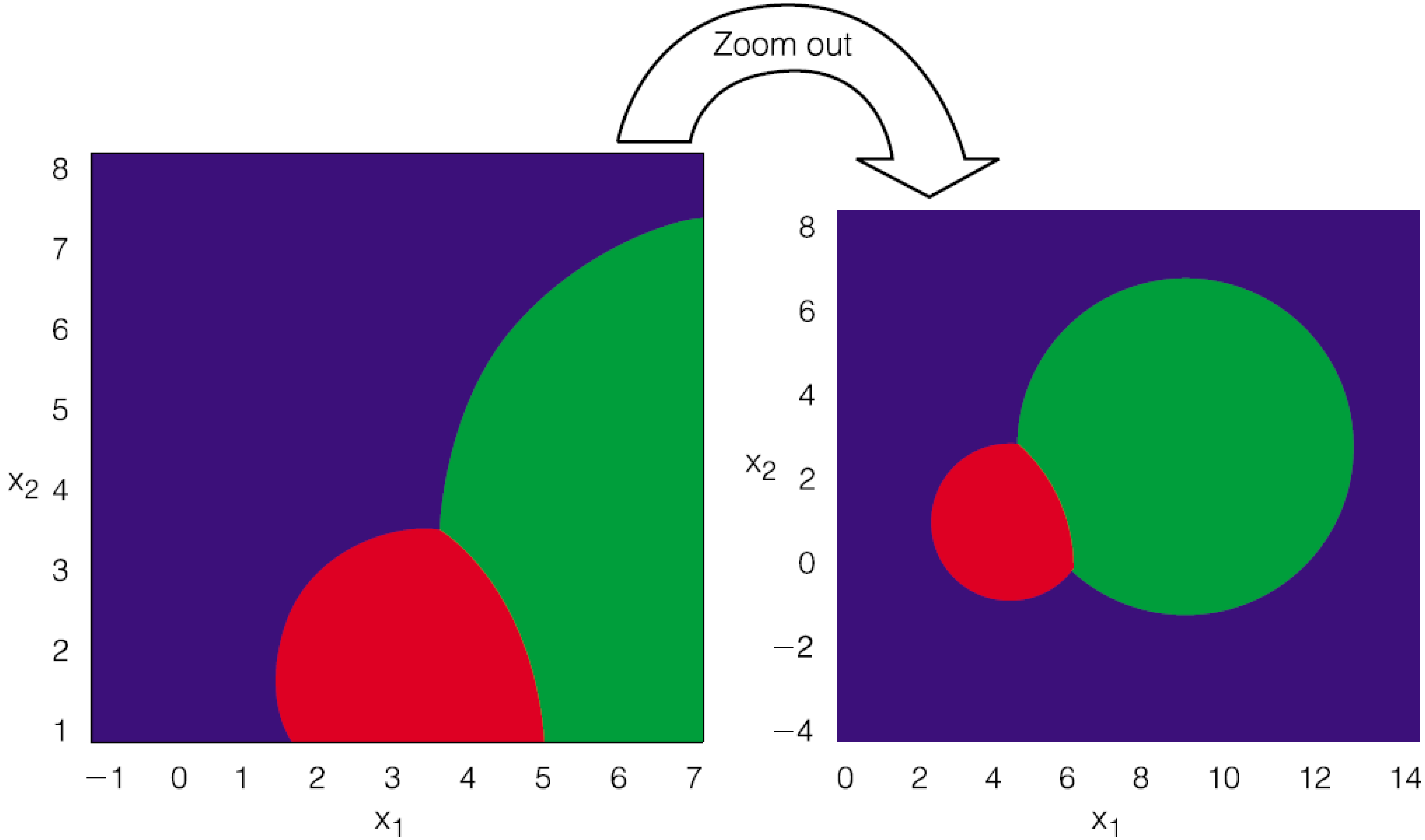
$$\Sigma_1 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$





❖ 데이터에 대한 통계적인 공분산 행렬의 종류

Case 4: $\mathbf{\Sigma_i = \sigma_i^2 I}$



[그림 6-10] Case 4의 특징 벡터의 분포와 결정 경계

❖ 데이터에 대한 통계적인 공분산 행렬의 종류

Case 5: $\Sigma_i \neq \Sigma_j$ (일반형)

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) - \frac{1}{2}\log(|\Sigma_i|) + \log(P(\omega_i))$$

$$g_i(x) = x^T W_i x + w_i^T x + w_{i0}$$

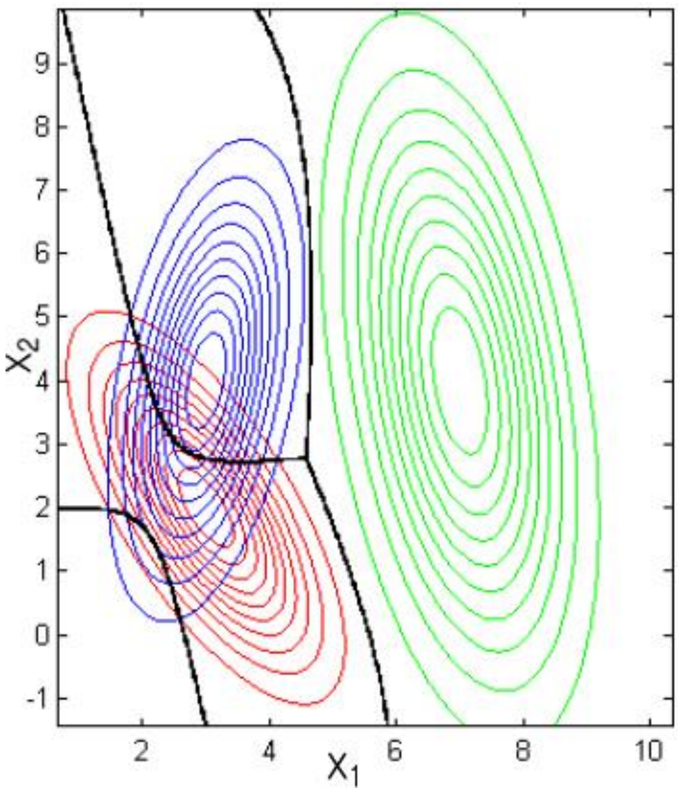
$$※ \begin{cases} W_i = -\frac{1}{2}\Sigma_i^{-1} \\ w_i = \Sigma_i^{-1}\mu_i \\ w_{i0} = -\frac{1}{2}\mu_i^T \Sigma_i^{-1}\mu_i - \frac{1}{2}\log\left(|\Sigma_i|\right) + \log(P(\omega_i)) \end{cases}$$

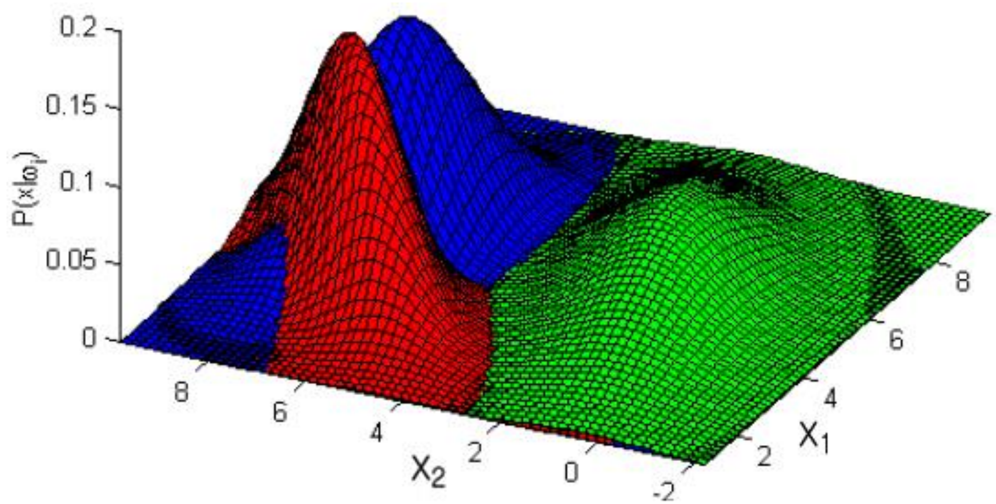
❖ 데이터에 대한 통계적인 공분산 행렬의 종류

Case 5: $\Sigma_i \neq \Sigma_j$ (일반형)

$$\mu_1 = [3 \quad 2]^T \qquad \mu_2 = [5 \quad 4]^T \qquad \mu_3 = [2 \quad 5]^T$$

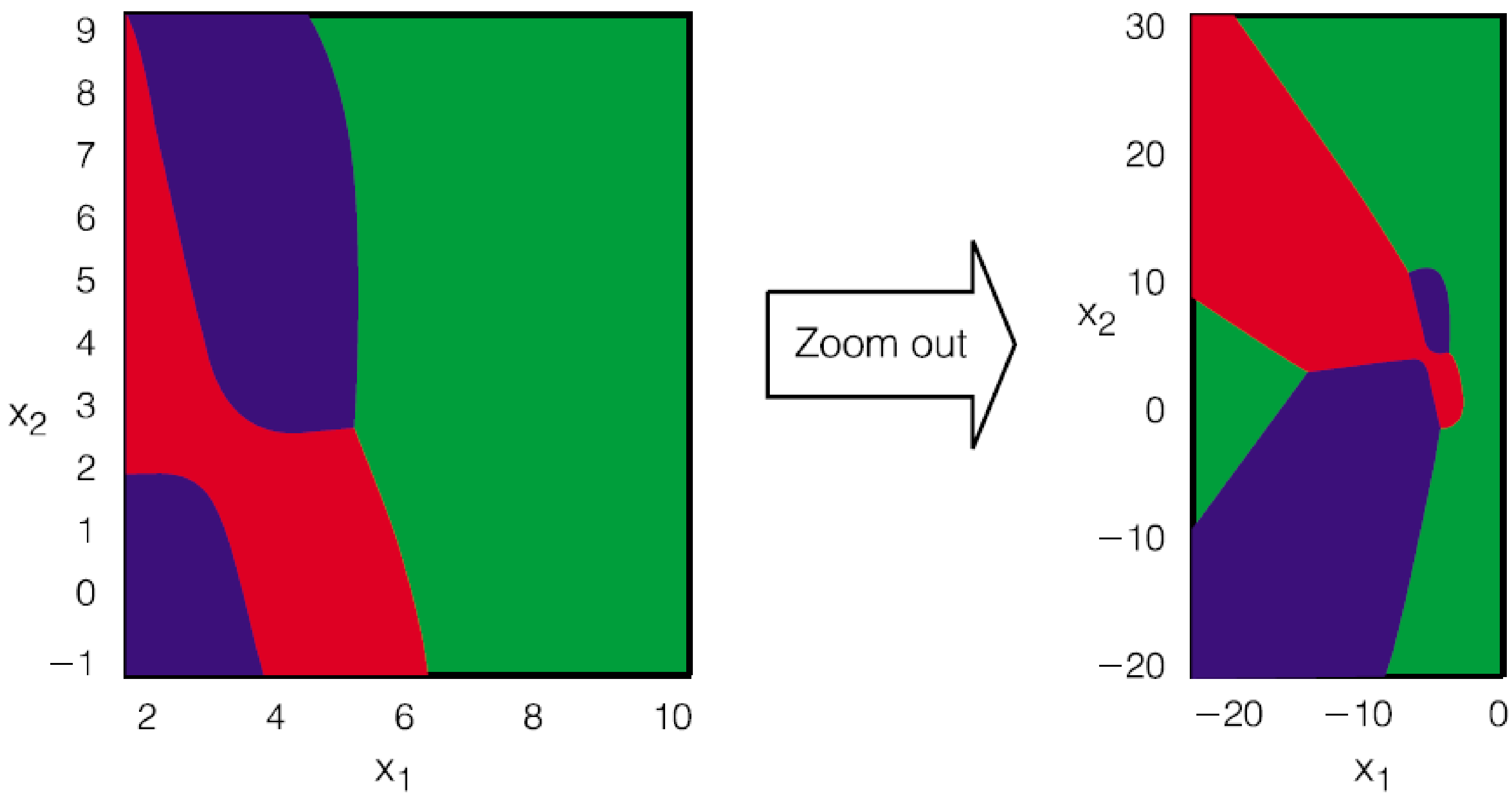
$$\Sigma_1 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} 1 & -1 \\ -1 & 7 \end{bmatrix} \qquad \Sigma_3 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 3 \end{bmatrix}$$





❖ 데이터에 대한 통계적인 공분산 행렬의 종류

Case 5: $\Sigma_i \neq \Sigma_j$ (일반형)



[그림 6-11] Case 5의 특징 벡터의 분포와 결정 경계

예제 **6-1**

3차원 특징 벡터 집합에서 정의되는 다음과 같은 가우시안 파라미터로 이루어진 클래스 $\boldsymbol{\omega}_1$, $\boldsymbol{\omega}_2$에 대한 선형 판별함수를 유도하고, 특징 벡터 $\mathbf{x}_u = \begin{bmatrix} 0.1 & 0.7 & 0.8 \end{bmatrix}^T$가 어느 클래스에 속하는지 결정하시오.

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T;\ \boldsymbol{\mu}_2 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T;\ \Sigma_1 = \Sigma_2 = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/4 \end{bmatrix};\ p(\boldsymbol{\omega}_2) = 2p(\boldsymbol{\omega}_1)$$

▷▷**풀이**

선형 판별함수 $g_1(\mathbf{x})$, $g_2(\mathbf{x})$는 다음과 같다.

$$g_i(\mathbf{x}) = -\frac{\mathbf{1}}{\mathbf{2\sigma^2}}(\mathbf{x}-\boldsymbol{\mu}_\mathrm{i})^T(\mathbf{x}-\boldsymbol{\mu}_\mathrm{i}) + \log \mathbf{P}(\boldsymbol{\omega}_\mathrm{i}) = -\frac{1}{2}\begin{bmatrix} \mathbf{x}-\boldsymbol{\mu}_\mathrm{x} \\ \mathbf{x}-\boldsymbol{\mu}_\mathrm{y} \\ \mathbf{x}-\boldsymbol{\mu}_\mathrm{z} \end{bmatrix}^T \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} \mathbf{x}-\boldsymbol{\mu}_\mathrm{x} \\ \mathbf{x}-\boldsymbol{\mu}_\mathrm{y} \\ \mathbf{x}-\boldsymbol{\mu}_\mathrm{z} \end{bmatrix} + \log \mathbf{P}(\boldsymbol{\omega}_\mathrm{i})$$

$$g_1(\mathbf{x}) = -\frac{1}{2}\begin{bmatrix}\mathbf{x}-0\\ \mathbf{x}-0\\ \mathbf{x}-0\end{bmatrix}^T\begin{bmatrix}4&0&0\\0&4&0\\0&0&4\end{bmatrix}\begin{bmatrix}\mathbf{x}-0\\ \mathbf{x}-0\\ \mathbf{x}-0\end{bmatrix} + \log\frac{1}{3};\quad g_2(\mathbf{x}) = -\frac{1}{2}\begin{bmatrix}\mathbf{x}-1\\ \mathbf{x}-1\\ \mathbf{x}-1\end{bmatrix}^T\begin{bmatrix}4&0&0\\0&4&0\\0&0&4\end{bmatrix}\begin{bmatrix}\mathbf{x}-1\\ \mathbf{x}-1\\ \mathbf{x}-1\end{bmatrix} + \log\frac{2}{3}$$

$$g_1(\mathbf{x}) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} g_2(\mathbf{x}) \Rightarrow -2(\mathbf{x}_1^2+\mathbf{x}_2^2+\mathbf{x}_3^2) + \log\frac{1}{3} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} -2((\mathbf{x}_1-1)^2+(\mathbf{x}_2-1)^2+(\mathbf{x}_3-1)^2) + \log\frac{2}{3}$$

$$\mathbf{x}_1+\mathbf{x}_2+\mathbf{x}_3 \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \frac{6-\log 2}{4} = 1.32$$

테스트 데이터 $\mathbf{x}_u = [0.1 \quad 0.7 \quad 0.8]^T$는 다음과 같다.

$$0.1+0.7+0.8 = 1.6 \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} 1.32 \Rightarrow \mathbf{x}_u \in \boldsymbol{\omega}_2$$