

## 2강. 패턴인식기 개발

## ※ 점검하기

Q1. 매트랩을 이용하여 다음 단계를 따라 간단한 분류 실험을 해 보시오.  
아래에 정의된 평균과 공분산을 가진 가우시안 분포를 따르는 두 클래스에 대하여 각각 25개의 학습 데이터와 25개의 테스트 데이터를 생성하여 2차원 평면상의 산점도가 어떻게 나오는지 설명하시오.

$$\mu_1 = \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \mu_2 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

## &lt;관련학습보기&gt;

## 2) 2차원 데이터 생성

$$p(x|C_1) \sim G(\mu_1, \Sigma_1)$$

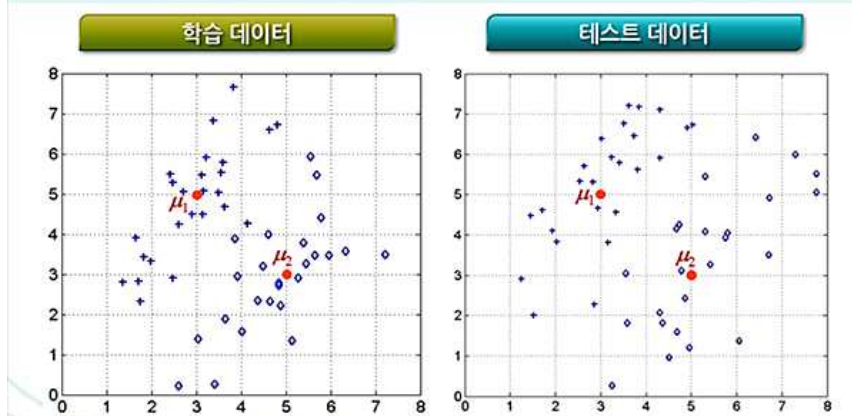
$$\mu_1 = \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

$$p(x|C_2) \sim G(\mu_2, \Sigma_2)$$

$$\mu_2 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

```
N=25;
m1= repmat([3,5], N, 1);
m2= repmat([5,3], N, 1);
s1=[1 1; 1 2];
s2=[1 1; 1 2];
X1=randn(N,2)*sqrtm(s1)+m1;
X2=randn(N,2)*sqrtm(s2)+m2;
plot(X1(:,1), X1(:,2), '+');
hold on;
plot(X2(:,1), X2(:,2), 'd');
save data2_1 X1 X2;
```

## 2) 2차원 데이터 생성



[참조] 1. 2차원 데이터의 분류의 「2) 2차원 데이터 생성」

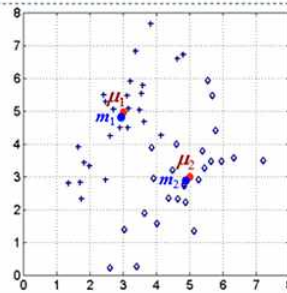
Q2. 학습 데이터를 이용하여 각 클래스의 평균과 공분산을 추정해 보고,  
(1)에서 정의한 실평균과 실공분산 값과 비교해 보시오.

<관련학습보기>

### 3) 학습 : 데이터의 분포 특성 분석

#### 1 데이터의 분포 특성 분석 → 결정경계

```
load data2_1;
m1 = mean(X1);
m2 = mean(X2);
s1 = cov(X1);
s2 = cov(X2);
save mean2_1 m1 m2 s1 s2;
```



평균과 공분산의 추정치

$$m_1 = \begin{pmatrix} 2.94 \\ 4.80 \end{pmatrix}, S_1 = \begin{pmatrix} 0.86 & 0.99 \\ 0.99 & 1.93 \end{pmatrix}, m_2 = \begin{pmatrix} 4.83 \\ 2.91 \end{pmatrix}, S_2 = \begin{pmatrix} 1.14 & 0.97 \\ 0.97 & 1.89 \end{pmatrix}$$

[참조] 1. 2차원 데이터의 분류의 「3) 학습 : 데이터의 분포 특성 분석」

Q3. (2)에서 계산한 평균을 이용하여 결정경계식을 찾아보고,  
그래프를 설명해 보시오.

<관련학습보기>

### 4) 분류 : 결정경계의 설정

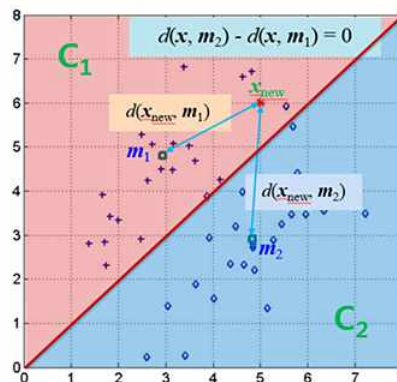
판별함수

$$g(x) = d(x, m_2) - d(x, m_1) = 0$$

$$\begin{aligned} d(x, m_2) - d(x, m_1) &= (x - m_1)^T (x - m_1) - (x - m_2)^T (x - m_2) \\ &= -2(m_1 - m_2)^T x + m_1^T m_1 - m_2^T m_2 \end{aligned}$$

클래스 라벨

$$y(x) = \begin{cases} 1 & \text{if } g(x) > 0 \\ -1 & \text{if } g(x) < 0 \end{cases}$$



[참조] 1. 2차원 데이터의 분류의 「4) 분류 : 결정경계의 설정」

Q4. (3)에서 찾은 결정경계를 이용하여 학습 데이터와 테스트 데이터에 대해 분류를 수행하고, 분류오차를 계산해 보시오.

<관련학습보기>

### 5) 성능 평가

#### 학습오차 계산 프로그램

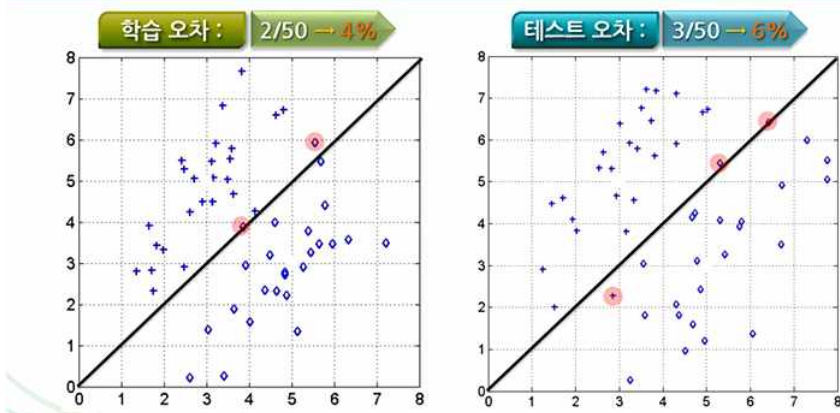
```
load data2_1;
load mean2_1;
Etrain=0;
N=size(X1,1);
for i=1:N
    d1=norm(X1(i,:) - m1);
    d2=norm(X1(i,:) - m2);
    if (d1-d2) > 0 Etrain = Etrain+1; end
    d1=norm(X2(i,:) - m1);
    d2=norm(X2(i,:) - m2);
    if (d1-d2) < 0 Etrain = Etrain+1; end
end
fprintf(1,'Training Error = %.3f\n', Etrain/50);
```

[참조] 1. 2차원 데이터의 분류의 「5) 성능 평가」

Q5. 테스트 데이터와는 별도로 충분히 많은 양의 데이터를 생성하여 일반화오차의 근사값을 계산해 보시오.

<관련학습보기>

### 5) 성능 평가



「2차원 데이터의 분류를 위한 패턴인식기 개발」의 각 단계  
[데이터 생성 ▶ 학습: 데이터의 분포 특성 분석 ▶ 분류: 결정경계의 설정 ▶ 성능 평가]에 따라 진행한다.  
따라서 교재/강의에서 사용한 매트랩 코드를 활용한다.

[참조] 1. 2차원 데이터의 분류의 「5) 성능 평가」

## ※ 정리하기

### 1. 패턴인식기의 개발 과정

- 1) 충분히 많은 양의 데이터를 수집(또는 생성)하여, 학습용 데이터와 테스트 데이터로 구분함
- 2) 필요한 경우 데이터에 대한 적절한 전처리를 수행하여 처리하기 좋은 형태의 데이터로 변환함  
→ 이 과정은 처리하는 데이터의 종류와 목적 등에 따라 달라짐
- 3) 학습 데이터를 이용하여 데이터의 분포 특성을 분석하고, 이로부터 데이터 분류 기준을 설정함
- 4) 설정된 분류 기준에 따라 학습 데이터와 테스트 데이터에 이용하여 성능을 평가함

### 2. 매트랩 프로그램

- 1) 교재에서 제시된 매트랩 프로그램([프로그램 2.1]~[프로그램 2.6])을 통해 매트랩의 기본적인 구문과 주요 함수들의 사용 방법을 보다 명확히 이해하고 활용 능력을 키우는 것이 중요함