

# 머신러닝과 빅데이터분석(R)

11주차 나이브베이즈

박길식 교수



고려사이버대학교  
THE CYBER UNIVERSITY OF KOREA



## 학습 목표



베이즈 이론을 설명할 수 있다.



베이즈 이론을 기반으로 베이지안 필터기를  
만들 수 있다.



# 학습 목차

1 베이지즈의 이해

2 베이지즈 실습

CHAPTER

# 베이지스의 이해

## 베이지 추정(Bayesian Estimation)

추론 대상의 사전 확률과 추가적인 정보를 기반으로  
해당 대상의 사후 확률을 추론하는 통계적 방법

### 예시

#### » 협력업체로부터 납품받은 기계의 성능을 평가 하는 경우

- 납품받은 몇 개의 부품을 무작위로 뽑아 이 표본에서 얻어진 정보만으로 모수(협력업체로부터 납품받은 기계 전체)의 성능을 해야 함
- 하지만 과거 납품 시 성능검사기록이나 비슷한 부품의 성능 자료, 이 부품의 물리적 특성에 관한 지식 등을 통해 이 부품의 사전 정보를 얻을 수 있음
- 이런 경우 단순히 표본을 통해 모수를 추정하기보다는 표본 정보와 사전 정보를 함께 사용하여 모수를 추정하는 것이 보다 바람직함

# [01] 베이지 추정

| Whether  | Play |
|----------|------|
| Sunny    | No   |
| Sunny    | No   |
| Overcast | Yes  |
| Rainy    | Yes  |
| Rainy    | Yes  |
| Rainy    | No   |
| Overcast | Yes  |
| Sunny    | No   |
| Sunny    | Yes  |
| Rainy    | Yes  |
| Sunny    | Yes  |
| Overcast | Yes  |
| Overcast | Yes  |
| Rainy    | No   |

Frequency Table

| Whether  | No | Yes |
|----------|----|-----|
| Overcast |    | 4   |
| Sunny    | 2  | 3   |
| Rainy    | 3  | 2   |
| Total    | 5  | 9   |

Likelihood Table 1

| Whether  | No    | Yes   |        |      |
|----------|-------|-------|--------|------|
| Overcast |       | 4     | = 4/14 | 0.29 |
| Sunny    | 2     | 3     | =5/14  | 0.36 |
| Rainy    | 3     | 2     | =5/14  | 0.36 |
| Total    | 5     | 9     |        |      |
|          | =5/14 | =9/14 |        |      |
|          | 0.36  | 0.64  |        |      |

Likelihood Table 2

| Whether  | No | Yes | Posterior Probability<br>for No | Posterior Probability<br>for Yes |
|----------|----|-----|---------------------------------|----------------------------------|
| Overcast |    | 4   | 0/5=0                           | 4/9=0.44                         |
| Sunny    | 2  | 3   | 2/5=0.4                         | 3/9=0.33                         |
| Rainy    | 3  | 2   | 3/5=0.6                         | 2/9=0.22                         |
| Total    | 5  | 9   |                                 |                                  |

## 01 베イズ 추정



날씨가 overcast일 때 경기를 할 확률

$$P(h|D) = \frac{P(D|h)P(h|D)}{P(D)}$$

$$P(\text{Yes}|\text{Overcast}) = P(\text{Overcast}|\text{Yes}) P(\text{Yes}) / P(\text{Overcast})$$

## [01] 베イズ 추정



### 날씨가 overcast일 때 경기를 할 확률

Likelihood Table 1

| Whether  | No    | Yes   |        |      |
|----------|-------|-------|--------|------|
| Overcast |       | 4     | = 4/14 | 0.29 |
| Sunny    | 2     | 3     | =5/14  | 0.36 |
| Rainy    | 3     | 2     | =5/14  | 0.36 |
| Total    | 5     | 9     |        |      |
|          | =5/14 | =9/14 |        |      |
|          | 0.36  | 0.64  |        |      |

Likelihood Table 2

| Whether  | No | Yes | Posterior Probability<br>for No | Posterior Probability<br>for Yes |
|----------|----|-----|---------------------------------|----------------------------------|
| Overcast |    | 4   | 0/5=0                           | 4/9=0.44                         |
| Sunny    | 2  | 3   | 2/5=0.4                         | 3/9=0.33                         |
| Rainy    | 3  | 2   | 3/5=0.6                         | 2/9=0.22                         |
| Total    | 5  | 9   |                                 |                                  |

#### 1 사전 확률

- $P(\text{Overcast}) = 4/14 = 0.29$
- $P(\text{Yes}) = 9/14 = 0.64$

#### 2 사후 확률

- $P(\text{Overcast}|\text{Yes}) = 4/9 = 0.44$



## [01] 베イズ 추정



날씨가 overcast일 때 경기를 할 확률

### 3 베イズ 정리 공식에 대입

$$\begin{aligned} P(\text{Yes}|\text{Overcast}) &= P(\text{Overcast}|\text{Yes}) P(\text{Yes}) / P(\text{Overcast}) \\ &= 0.44 * 0.64 / 0.29 = 0.98 \end{aligned}$$

날씨가 Overcast일 때 축구를 할 확률 → 0.98

## 01 베이지 추정

 날씨가 overcast일 때 경기를 하지 않을 확률

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

$$P(\text{No}|\text{Overcast}) = P(\text{Overcast}|\text{No}) P(\text{No}) / P(\text{Overcast})$$

## [01] 베イズ 추정

 날씨가 overcast일 때 경기를 하지 않을 확률

Likelihood Table 1

| Whether  | No    | Yes   |        |      |
|----------|-------|-------|--------|------|
| Overcast |       | 4     | = 4/14 | 0.29 |
| Sunny    | 2     | 3     | =5/14  | 0.36 |
| Rainy    | 3     | 2     | =5/14  | 0.36 |
| Total    | 5     | 9     |        |      |
|          | =5/14 | =9/14 |        |      |
|          | 0.36  | 0.64  |        |      |

Likelihood Table 2

| Whether  | No | Yes | Posterior Probability<br>for No | Posterior Probability<br>for Yes |
|----------|----|-----|---------------------------------|----------------------------------|
| Overcast |    | 4   | 0/5=0                           | 4/9=0.44                         |
| Sunny    | 2  | 3   | 2/5=0.4                         | 3/9=0.33                         |
| Rainy    | 3  | 2   | 3/5=0.6                         | 2/9=0.22                         |
| Total    | 5  | 9   |                                 |                                  |

### 1 사전 확률

- $P(\text{Overcast}) = 4/14 = 0.29$
- $P(\text{No}) = 5/14 = 0.36$

### 2 사후 확률

- $P(\text{Overcast}|\text{No}) = 0/5 = 0$

## [01] 베イズ 추정

 날씨가 overcast일 때 경기를 하지 않을 확률

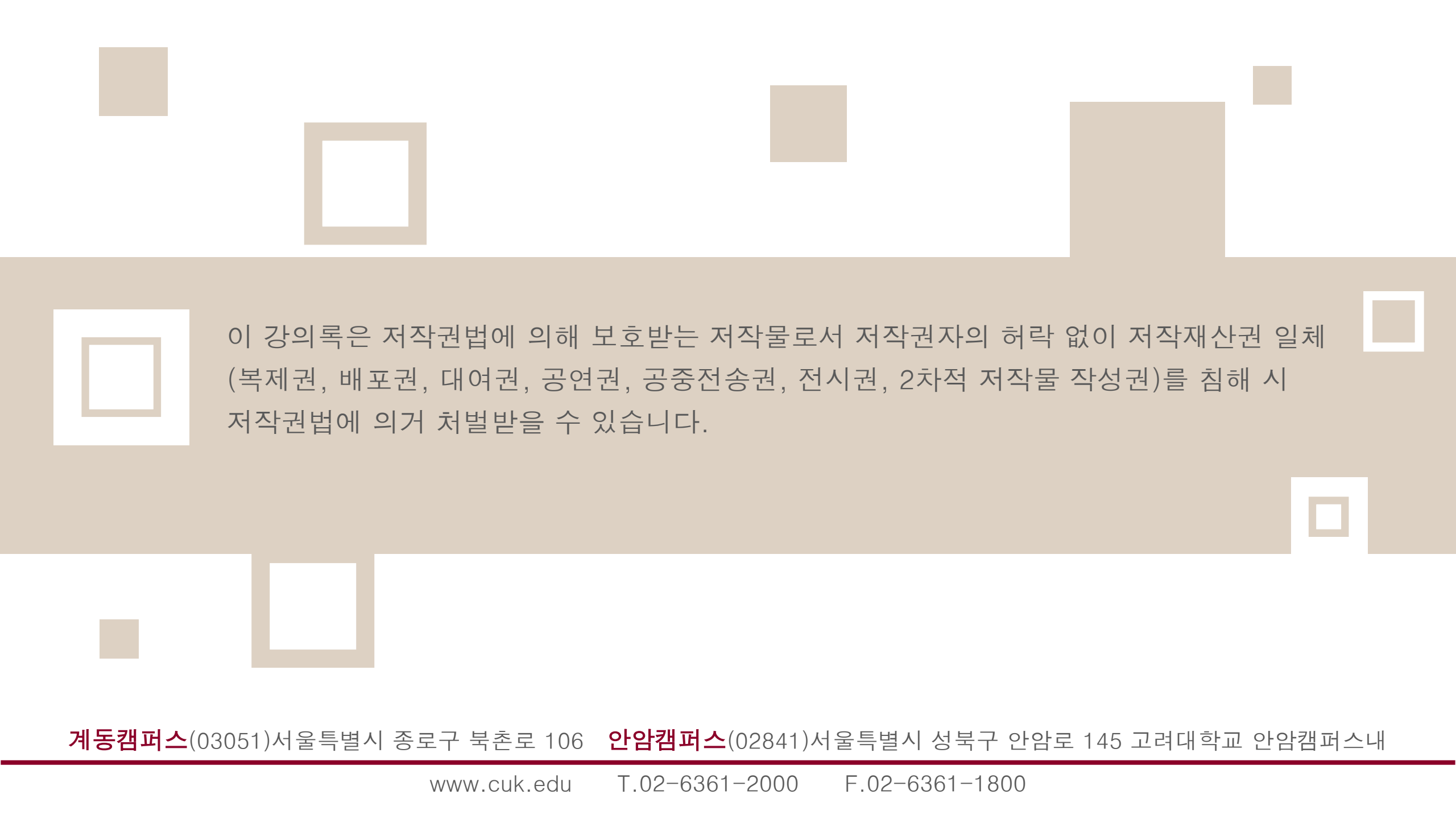
### 3 베イズ 정리 공식에 대입

$$\begin{aligned} P(\text{No}|\text{Overcast}) &= P(\text{Overcast}|\text{No}) P(\text{No}) / P(\text{Overcast}) \\ &= 0 * 0.36 / 0.29 = 0 \end{aligned}$$

날씨가 Overcast일 때 축구를 할 확률

$$P(\text{Yes}|\text{Overcast}) = 0.98, P(\text{No}|\text{Overcast}) = 0$$

- 날씨가 Overcast일 때 축구를 하는 확률은 0.98, 축구를 하지 않을 확률은 0
- 두 확률을 비교한 뒤 더 높은 확률의 Label로 분류



이 강의록은 저작권법에 의해 보호받는 저작물로서 저작권자의 허락 없이 저작권재산권 일체 (복제권, 배포권, 대여권, 공연권, 공중전송권, 전시권, 2차적 저작물 작성권)를 침해 시 저작권법에 의거 처벌받을 수 있습니다.