기계학습원론 HW5

2021312738 소프트웨어학과 김서환

모든 문제에서 naïve Bayesian classifier의 가정을 참고하여 class(yes/no)가 주어졌을 때, 각 Input들(Outlook과 같은 속성)은 서로 independent하다고 가정했습니다. 그리고 Temp 값은 숫자형 데이터로 Gaussian distribution을 따른다고 가정했습니다.

또한, 수동으로 계산이 힘든 점을 고려해서 yes/no 클래스 별 데이터의 Temp값의 평균, 표준편차와 P(T=25|yes), P(T=25|no), P(T=20|yes), P(T=20|no), P(T=24|yes), P(T=24|no)의 확률을 파이썬 프로그램을 통해 구했습니다.

그리고 S-wind, W-wind의 확률(1번 발생할 확률)은 수동으로 yes/no 클래스 별 S-wind의 개수/전체 wind의 개수, W-wind의 개수/전체 wind의 개수로 구했습니다.

```python
import math

def calc_mean_std(data):
    n = len(data)
    mean = sum(data) / n
    variance = sum((x - mean) ** 2 for x in data) / (n-1)
    std_dev = math.sqrt(variance)
    return mean, std_dev

def calc_probability(x, mu, sigma):
    exponent = math.exp(-((x - mu) ** 2) / (2 * sigma ** 2))
    probability = (1 / (math.sqrt(2 * math.pi) * sigma)) * exponent
    return probability

nodata = [30,31,13,25,20,18]
yesdata = [28,23,10,14,15,22,19,33]

yesmean, yesstd = calc_mean_std(yesdata)
nomean, nostd = calc_mean_std(nodata)

problem1y = calc_probability(25,yesmean,yesstd)
problem1n = calc_probability(25,nomean,nostd)
problem2y = calc_probability(20,yesmean,yesstd)
problem2n = calc_probability(20,nomean,nostd)
problem3y = calc_probability(24,yesmean,yesstd)
problem3n = calc_probability(24,nomean,nostd)

print("problem1 Temp yes probability : ", problem1y)
print("problem1 Temp no probability : ", problem1n)
print("problem2 Temp yes probability : ", problem2y)
print("problem2 Temp no probability : ", problem2n)
print("problem3 Temp yes probability : ", problem3y)
print("problem3 Temp no probability : ", problem3n)
```

```
C:\Users\kksh3\OneDrive\바탕 화면\VSCODE\Python>python test.py
problem1 Temp yes probability :  0.043992812794088225
problem1 Temp no probability :   0.053750494404141744
problem2 Temp yes probability :  0.05227091344534292
problem2 Temp no probability :   0.051994101207002215
problem3 Temp yes probability :   0.04713386723608115
problem3 Temp no probability :   0.05556621965994455
```

1.Overcast, T=25, Normal, s-wind=3, w-wind=2 => **yes**

1. overcast, T=25, Normal, s-wind=3, w-wind=2 → yes와 no의 비교

$h_1 = P(yes \mid overcast, T=25, Normal, SW=3, WW=2)$ 라하자

$$= \frac{P(overcast, T=25, Normal, SW=3, WW=2 \mid yes)\, P(yes)}{P(overcast, T=25, Normal, SW=3, WW=2)}$$

$$= \frac{P(overcast \mid yes)\, P(T=25 \mid yes)\, P(Normal \mid yes)\, P(SW=3, WW=2 \mid yes)\, P(yes)}{P(overcast, T=25, Normal, SW=3, WW=2)}$$

$$\Rightarrow P(yes) = \frac{8}{14}$$

전부
공통 $P(overcast \mid yes) = \frac{3}{8}$ 소수점
(분자) $P(T=25 \mid yes) = 0.044$ ← 3 자리까지 표현
$P(Normal \mid yes) = \frac{6}{8}$
$P(SW=3, WW=2 \mid yes) = {}_5C_3 \times \left(\frac{15}{35}\right)^3 \times \left(\frac{20}{35}\right)^2$

<wind의 확률>

	SW	WW
No	$\frac{16}{31}$	$\frac{15}{31}$
Yes	$\frac{15}{35}$	$\frac{20}{35}$

$P(overcast, T=25, Normal, SW=3, WW=2) = \alpha$ 라고 가정 (분모)

$$\therefore h_1 = \frac{0.0018\cdots}{\alpha}$$

$h_2 = P(no \mid overcast, T=25, Normal, SW=3, WW=2)$ 라하자

$$= \frac{P(overcast, T=25, Normal, SW=3, WW=2 \mid no)\, P(no)}{P(overcast, T=25, Normal, SW=3, WW=2)}$$

$$= \frac{P(overcast \mid no)\, P(T=25 \mid no)\, P(Normal \mid no)\, P(SW=3, WW=2 \mid no)\, P(no)}{P(overcast, T=25, Normal, SW=3, WW=2)}$$

$P(no) = \frac{6}{14}$ $P(overcast \mid no) = \frac{1}{6}$ $P(T=25 \mid no) = 0.054$ $P(Normal \mid no) = \frac{1}{6}$

$P(SW=3, WW=2 \mid no) = {}_5C_3 \times \left(\frac{16}{31}\right)^3 \times \left(\frac{15}{31}\right)^2$, 분모 $= \alpha$ (yes의 경우와 같음)

$$\therefore h_2 = \frac{0.000207\cdots}{\alpha}$$

$\therefore h_1 > h_2$ 이므로 **yes**로 분류됨!

2.Rain, T=20, High, s-wind=0, w-wind=3 => **<u>no</u>**

2. Rain, $T=20$, High, $SW=0, WW=3$

1번문제과 같은 과정을 통해 k_1 (yes일때확률), k_2 (no일때확률)를 구해 비교해보겠습니다.

⟨wind의 확률⟩

$k_1 = \dfrac{P(yes)\,P(Rain|yes)\,P(T=20|yes)\,P(High|yes)\,P(sw=0,ww=3|yes)}{P(Rain, T=20, High, sw=0, ww=3) = \alpha 가치 o}$

	SW	WW
No	$16/31$	$15/31$
Yes	$15/35$	$20/35$

$P(yes) = \dfrac{8}{14}$

$P(Rain|yes) = \dfrac{3}{8}$

$P(T=20|yes) = 0.052$

$P(High|yes) = \dfrac{2}{8}$

$P(sw=0, ww=3|yes) = {}_3C_3 \times \left(\dfrac{20}{35}\right)^3 \Rightarrow k_1 = \dfrac{0.00059\cdots}{\alpha}$

$k_2 = \dfrac{P(no)\,P(Rain|no)\,P(T=20|no)\,P(High|no)\,P(sw=0,ww=3|no)}{P(Rain, T=20, High, sw=0, ww=3) = \alpha 가치 o}$

$P(no) = \dfrac{6}{14}$

$P(Rain|no) = \dfrac{2}{6}$

$P(T=20|no) = 0.052$

$P(High|no) = \dfrac{5}{6}$

$P(sw=0, ww=3|no) = {}_3C_3 \times \left(\dfrac{15}{31}\right)^3$

$\Rightarrow k_2 = \dfrac{0.00070\cdots}{\alpha}$

$\therefore k_1 < k_2$ 이므로 $\boxed{No}$ 로 분류됨!

3.Sunny, T=24, Normal, s-wind=4, w-wind=4 => **yes**

3. Sunny, T=24, Normal, SW=4, WW=4

1번문제과 같은 과정을 통해 h_1 (yes일때확률), h_2 (no일때확률)를 구해 비교해보겠습니다.

$h_1 = \dfrac{P(yes)\,P(sunny\,|\,yes)\,P(T=24\,|\,yes)\,P(Normal\,|\,yes)\,P(sw=4,\,WW=4\,|\,yes)}{P(sunny,\,T=24,\,Normal,\,sw=4,\,ww=4) = \alpha 가지}$

⟨wind의 확률⟩

	SW	WW
No	$16/31$	$15/31$
Yes	$15/35$	$20/35$

$P(yes) = 8/14$

$P(sunny\,|\,yes) = 2/8$

$P(T=24\,|\,yes) = 0.047$

$P(Normal\,|\,yes) = 6/8$

$P(sw=4,\,ww=4\,|\,yes) = {}_8C_4 \times \left(\dfrac{15}{35}\right)^4 \times \left(\dfrac{20}{35}\right)^4$

$\Rightarrow h_1 = \dfrac{0.00126\cdots}{\alpha}$

$h_2 = \dfrac{P(no)\,P(sunny\,|\,no)\,P(T=24\,|\,no)\,P(Normal\,|\,no)\,P(sw=4,\,ww=4\,|\,no)}{P(sunny,\,T=24,\,Normal,\,sw=4,\,ww=4) = \alpha 가지}$

$P(no) = 6/14$

$P(sunny\,|\,no) = 3/6$

$P(T=24\,|\,no) = 0.056$

$P(Normal\,|\,no) = 1/6$

$P(sw=4,\,ww=4\,|\,no) = {}_8C_4 \times \left(\dfrac{16}{31}\right)^4 \times \left(\dfrac{15}{31}\right)^4$

$\Rightarrow h_2 = \dfrac{0.00036\cdots}{\alpha}$

$\therefore h_1 > h_2$ 이므로 **yes**로 분류!