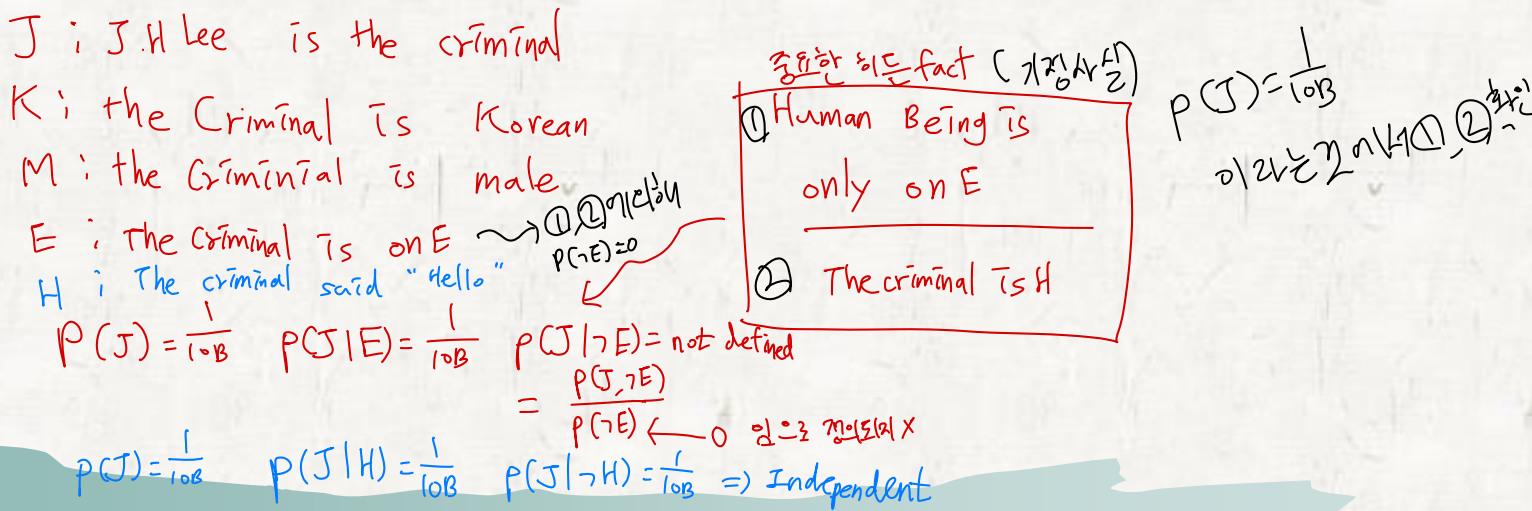


Naïve Bayesian Classifier

A Simple Probabilistic Approach

Recap: Probability

- Conditional Probability: 조건부 확률

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \rightarrow P(A,B)$$

- Probability of A given B
- Probability of A with assumption that B has happened
- Probability of A evaluated with only the samples in B
- Probability of A evaluated with only the samples having B

- Probabilistic Independence

If and only if $P(A,B) = P(A)P(B)$ or $P(A|B) = P(A|\neg B) = P(A)$

- Bayesian Rule

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

	X1	X2
1	a1	b2
2	a2	b3
3	a1	b1
4	a3	b2
5	a2	b2
6	a2	b1
7	a3	b3
8	a2	b2

Recap: Probability

- Conditional Probability, Independence & Bayesian Rule
  - A and B are independent given C
    - $P(A, B|C) = P(A|C)\ P(B|C)$
    - $P(A|B, C) = P(A|\neg B, C) = P(A|C)$
  - Bayesian Rule given C

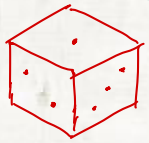
$$P(A|B, C) = \frac{P(A, B|C)}{P(B|C)} = \frac{P(B|A, C)P(A|C)}{P(B|C)}$$

  - Which are True?
    - If A and B are indep., A and B are indep. given C
    - If A and B are indep. given C, A and B are indep.
    - If A and B are indep. given C, A and B are indep. Given D

Recap: Probability

1~3 page 중요하다!

- Multinomial Distribution



  - Die tossing with not fair Dice

$$P(1)=\frac{1}{12}, P(2)=\frac{1}{12}, P(3)=\frac{1}{12}, P(4)=\frac{3}{12}, P(5)=\frac{3}{12}, P(6)=\frac{3}{12}$$

    What is the probability of "1"=3, "2"=1, "4"=2 ← 6번뽑

    순서에 대한 조합

$$C(6,3)\times C(3,1)\times C(2,2)\times\left(\frac{1}{12}\right)^3\times\left(\frac{1}{12}\right)\times\left(\frac{3}{12}\right)^2$$

  - You toss a die several times, and repeat this several times

Die face	1	2	3	4	5	6
Trial 1	1	2	2	2	2	1
Trial 2	2	1	0	1	2	0
Trial 3	2	2	1	1	1	2

10 times, 6 times, 9 times → total 25번

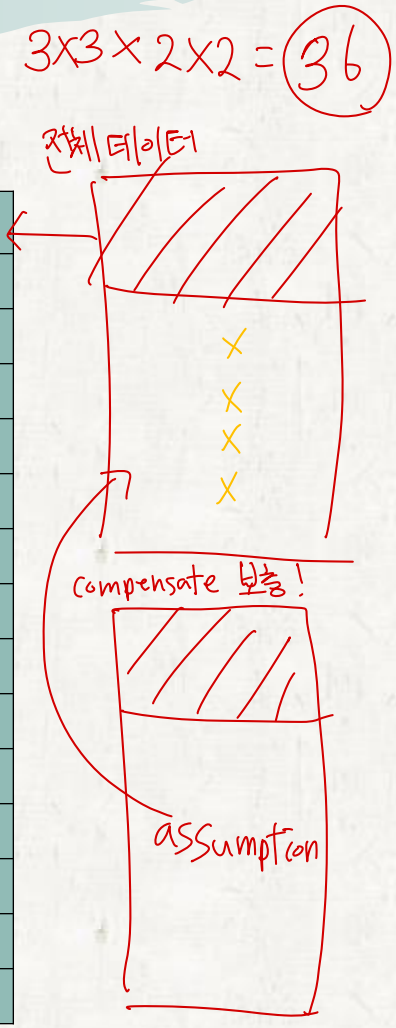
    - Estimate $P(1), P(2), \cdots, P(6)$

$= \frac{5}{25}$ $= \frac{5}{25}$ $\cdots$ $= \frac{3}{25}$

Tennis Playing

- Tennis playing record

	Outlook	Temp	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No



Tennis Playing

- Today is Sunny, Mild, High and Strong
  - Will he play tennis today?

- How can you predict his tennis playing?
  - Let's use Probability

$$P(yes \mid sunny, mild, high, strong)$$
$$P(no \mid sunny, mild, high, strong)$$

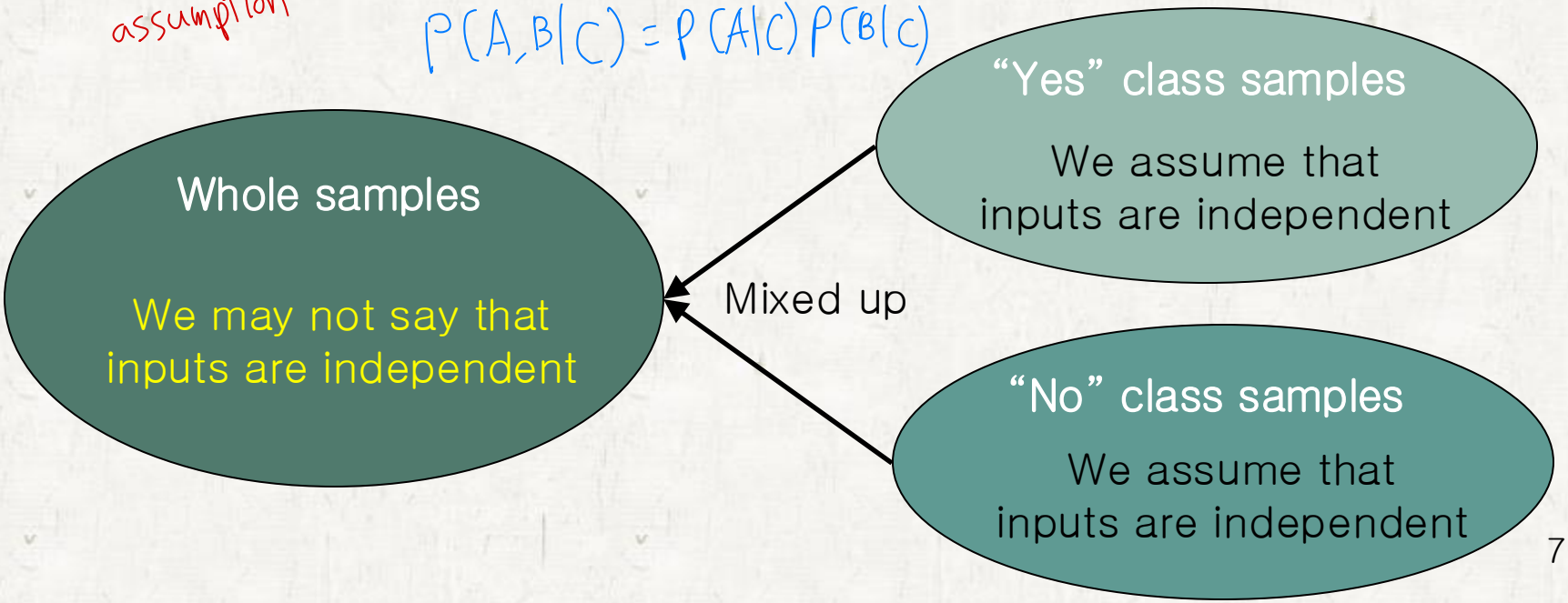
  - However, we need to observe a lot of "$sunny, mild, high, strong$" cases

Tennis Playing

- Naïve Bayesian Classifier
  - We assume the strongest assumption on independence to compensate shortage of data samples

If class is given,
Inputs are independent from each other.

$P(A, B|C) = P(A|C)P(B|C)$



Tennis Playing

- Naïve Bayesian Assumption

	A	B	Class
1	O	O	Yes
2	O	X	Yes
3	X	O	Yes
4	X	X	Yes

$P(A|Yes) = 0.5$

$P(B|Yes) = 0.5$

$P(A,B|Yes) = 0.25$

$= P(A)P(B)$ Independent!

$P(A|No) = 0.75$

$P(B|No) = 0.75$

$P(A,B|No) = 0.5625$

Independent !

	A	B	Class
1	O	O	No
2	O	O	No
3	O	O	No
4	O	O	No
5	O	O	No
6	O	O	No
7	O	O	No
8	O	O	No
9	O	O	No
10	O	X	No
11	O	X	No
12	O	X	No
13	X	O	No
14	X	O	No
15	X	O	No
16	X	X	No

아무것도 주어지지 X

dependent

$P(A) = 0.7$ $\frac{14}{20}$

$P(B) = 0.7$ $\frac{14}{20}$

$P(A,B) = 0.5$

$\neq P(A)P(B)$

	A	B	Class
1	O	O	No
2	O	O	No
3	O	O	No
4	O	O	No
5	O	O	No
6	O	O	No
7	O	O	No
8	O	O	No
9	O	O	No
10	O	X	No
11	O	X	No
12	O	X	No
13	X	O	No
14	X	O	No
15	X	O	No
16	X	X	No
17	O	O	Yes
18	O	X	Yes
19	X	O	Yes
20	X	X	Yes

yes

Tennis Playing

- Naïve Bayesian Classifier

switch!! 이해하보기

Bayesian Rule

$$P(yes|sunny,mild,high,strong) = \frac{P(sunny,mild,high,strong|yes)P(yes)}{P(sunny,mild,high,strong)}$$

독립 가정 : $P(A,B|C)=P(A|C)P(B|C)$ 이용

$$= \frac{P(sunny|yes)P(mild|yes)P(high|yes)P(strong|yes)P(yes)}{P(sunny,mild,high,strong)}$$

각각을 표에서 구함

$$P(no|sunny,mild,high,strong) = \frac{P(sunny,mild,high,strong|no)P(no)}{P(sunny,mild,high,strong)}$$

$$= \frac{P(sunny|no)P(mild|no)P(high|no)P(strong|no)P(no)}{P(sunny,mild,high,strong)}$$

Tennis Playing

- Naïve Bayesian Classifier
  - Then, how to obtain the probabilities?

$P(yes) = ?$

$P(no) = ?$

$P(sunny|yes) = ?$

$P(mild|yes) = ?$

$P(high|yes) = ?$

$P(strong|yes) = ?$

$P(sunny|no) = ?$

$P(mild|no) = ?$

$P(high|no) = ?$

$P(strong|no) = ?$

$$P(yes|sunny, mild, high, strong) = \frac{P(sunny|yes)P(mild|yes)P(high|yes)P(strong|yes)P(yes)}{P(sunny, mild, high, strong)}$$

$$P(no|sunny, mild, high, strong) = \frac{P(sunny|no)P(mild|no)P(high|no)P(strong|no)P(no)}{P(sunny, mild, high, strong)}$$

Tennis Playing

- Naïve Bayesian Classifier
  - Let's ESTIMATE from the table

$P(yes) = 9/14$

$P(no) = 5/14$

$P(sunny|yes) = 2/9$

$P(mild|yes) = 4/9$

$P(high|yes) = 3/9$

$P(strong|yes) = 3/9$

$P(sunny|no) = 3/5$

$P(mild|no) = 2/5$

$P(high|no) = 4/5$

$P(strong|no) = 3/5$

	Outlook	Temp	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

Tennis Playing

- Naïve Bayesian Classifier

$$P(yes \mid sunny, mild, high, strong)$$

$$= \frac{P(yes)P(sunny \mid yes)P(mild \mid yes)P(high \mid yes)P(strong \mid yes)}{P(sunny, mild, high, strong)}$$

given! yes, no가 주어졌을때 ↓ 독립이지(전처리됐)

$\neq P(s)\,P(m)\,P(h)\,P(s)$

$$= \frac{1}{\alpha}\left(\frac{9}{14}\times\frac{2}{9}\times\frac{4}{9}\times\frac{3}{9}\times\frac{3}{9}\right) = \frac{0.007055}{\alpha}$$

$$P(sunny, mild, high, strong) = \alpha$$

가정

$\Rightarrow$ No! 값을 비교는 할 수있다.

$$P(no \mid sunny, mild, high, strong)$$

$$= \frac{P(no)P(sunny \mid no)P(mild \mid no)P(high \mid no)P(strong \mid no)}{P(sunny, mild, high, strong)}$$

$$= \frac{1}{\alpha}\left(\frac{5}{14}\times\frac{3}{5}\times\frac{2}{5}\times\frac{4}{5}\times\frac{3}{5}\right) = \frac{0.04114}{\alpha}$$

둘이 더해서 α값은 구해도 되는데 의미없다.

Will he play or not?

Discussion

"Some Special case"

- Let's do it once more with the following data

	Outlook	Temp	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Rain	Mild	High	Weak	Yes
4	Rain	Cool	Normal	Strong	No
5	Sunny	Mild	High	Weak	No
6	Rain	Mild	Normal	Weak	Yes
7	Overcast	Mild	High	Strong	Yes
8	Rain	Mild	High	Strong	No

- Today is Sunny, Mild, High and Strong
  - Will he play tennis today?

Discussion

- Naïve Bayesian Classifier
  - Let's ESTIMATE the followings, but HOW?

이런 데이터에 smoothing을 적용하는 건 바람직!

$$P(yes) = 3/8$$
$$P(no) = 5/8$$
$$P(sunny \mid yes) = 0/3$$
$$P(mild \mid yes) = 3/3$$
$$P(high \mid yes) = 2/3$$
$$P(strong \mid yes) = 1/3$$
$$P(sunny \mid no) = 3/5$$
$$P(mild \mid no) = 2/5$$
$$P(high \mid no) = 4/5$$
$$P(strong \mid no) = 3/5$$

small size dataset!

	Outlook	Temp	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Rain	Mild	High	Weak	Yes
4	Rain	Cool	Normal	Strong	No
5	Sunny	Mild	High	Weak	No
6	Rain	Mild	Normal	Weak	Yes
7	Overcast	Mild	High	Strong	Yes
8	Rain	Mild	High	Strong	No

해결방안 Sample 추가 Based on Expert Knowledge

You!!

??

Discussion

- Naïve Bayesian Classifier

$$P(yes \mid sunny, mild, high, strong)$$

$$= \frac{P(yes)P(sunny \mid yes)P(mild \mid yes)P(high \mid yes)P(strong \mid yes)}{P(sunny, mild, high, strong)}$$

$$= \frac{1}{\alpha}\left(\frac{3}{8} \times \frac{0}{3} \times \frac{3}{3} \times \frac{2}{3} \times \frac{1}{3}\right) = 0$$

$$P(no \mid sunny, mild, high, strong)$$

$$= \frac{P(no)P(sunny \mid no)P(mild \mid no)P(high \mid no)P(strong \mid no)}{P(sunny, mild, high, strong)}$$

$$= \frac{1}{\alpha}\left(\frac{5}{8} \times \frac{3}{5} \times \frac{2}{5} \times \frac{4}{5} \times \frac{3}{5}\right) = \frac{0.072}{\alpha}$$

Is it OK?

Discussion

- Smoothing
  - Adjust the probability

$$P(x_i = t|y = c) = \frac{N_{x_i=t,y=c} + \alpha}{N_{y=c} + \alpha n_i}$$

$P(\text{sunny}|\text{yes}) = \frac{N_{\text{sunny}} + \alpha \cdot 1}{N_{\text{yes}} + \alpha \cdot 10}$ (α: 내가 추가한 샘플이 자신있다; α↑, $\frac{1}{10}$, number of)

- $N_{x_i=t,y=c}$: number of training examples for which $x_i = t, y = c$
- $N_{y=c}$: number of examples for which $y = c$
- n_i : number of values x_i can take
- α : smoothing parameter ($\alpha \geq 0$)
  - $\alpha = 1$: Laplace smoothing
  - $\alpha < 1$: Lidstone smoothing

추가하는 작업 ; smoothing

↳ avoid overfitting하는 좋은 방법

과소적합 → 너무 단순한모델, 학습부족으로 패턴 제대로 학습되지 못한상태

underfitting

특정데이터에만과도한 신뢰 ; 훈련데이터에 없는 데이터를 무시해버림

(No sample) too much follow Data

↑

overfitting ; smoothing 하기 이전의 과적합 상태임 !

Discussion

- Smoothing

Observed samples from given data

$$P(x_i = t|y = c) = \frac{N_{x_i=t,y=c} + \alpha}{N_{y=c} + \alpha n_i}$$

Implicit observations by Expert Knowledge

  - If there are more αn_i samples of which y is c, we may observe more α samples of which x_i is t and y is c

$$P(sunny|yes) = \frac{0+1}{3+2} = \frac{1}{5}$$

NB with Real Values

- Tennis playing record with real values

	Outlook	Temp(°C)	Humidity	Wind	Play
1	Sunny	30	High	Weak	No
2	Sunny	31	High	Strong	No
3	Overcast	28	High	Weak	Yes
4	Rain	23	High	Weak	Yes
5	Rain	10	Normal	Weak	Yes
6	Rain	13	Normal	Strong	No
7	Overcast	14	Normal	Strong	Yes
8	Sunny	25	High	Weak	No
9	Sunny	15	Normal	Weak	Yes
10	Rain	22	Normal	Weak	Yes
11	Sunny	19	Normal	Strong	Yes
12	Overcast	20	High	Strong	Yes
13	Overcast	33	Normal	Weak	Yes
14	Rain	18	High	Strong	No

NB with Real Values

- Today is
  - Outlook=Sunny, Temp=22, Humidity=High and Wind=Strong

- Will he play tennis today?

$$P(Yes|O = sunny, T = 22, H = High, W = Strong)$$

$$= \frac{P(sunny\,|\,yes)P(T = 22\,|\,yes)P(high\,|\,yes)P(strong|yes)P(yes)}{P(sunny, mild, high, strong)}$$

$$P(No|O = sunny, T = 22, H = High, W = Strong)$$

$$= \frac{P(sunny\,|\,no)P(T = 22\,|\,no)P(high\,|\,\mathrm{no})P(strong|no)P(no)}{P(sunny, mild, high, strong)}$$

NB with Real Values

- If there are continuous values, how can I ESTIMATE the probability?

$P(yes) = 9/14$

$P(no) = 5/14$

$P(sunny|yes) = 2/9$

$P(T = 22|yes) = ??$

$P(high|yes) = 3/9$

$P(strong|yes) = 3/9$

$P(sunny|no) = 3/5$

$P(T = 22|no) = ??$

$P(high|no) = 4/5$

$P(strong|no) = 3/5$

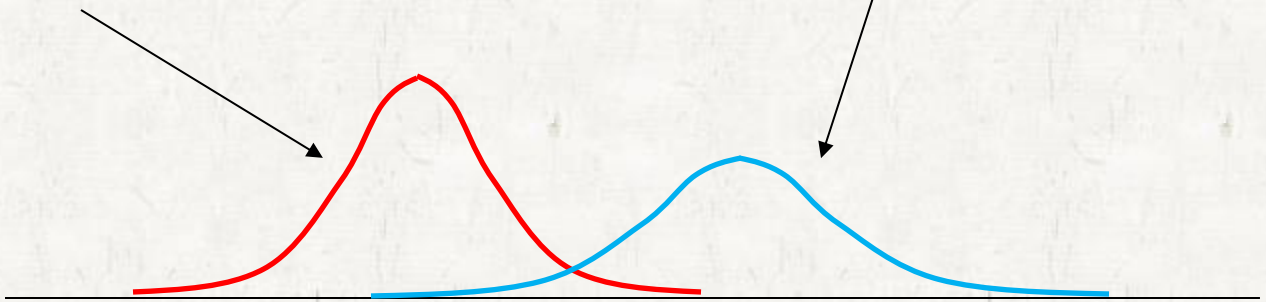
	Outlook	Temp	Humidity	Wind	Play
1	Sunny	30	High	Weak	No
2	Sunny	31	High	Strong	No
3	Overcast	28	High	Weak	Yes
4	Rain	23	High	Weak	Yes
5	Rain	10	Normal	Weak	Yes
6	Rain	13	Normal	Strong	No
7	Overcast	14	Normal	Strong	Yes
8	Sunny	25	High	Weak	No
9	Sunny	15	Normal	Weak	Yes
10	Rain	22	Normal	Weak	Yes
11	Sunny	19	Normal	Strong	Yes
12	Overcast	20	High	Strong	Yes
13	Overcast	33	Normal	Weak	Yes
14	Rain	18	High	Strong	No

NB with Real Values

- Measurement values -> modeled with the Gaussian dist.
  - Assumption: Temperature of each class will follow the Gaussian distribution

	Outlook	Temp	Humidity	Wind	Play
3	Overcast	28	High	Weak	Yes
4	Rain	23	High	Weak	Yes
5	Rain	10	Normal	Weak	Yes
7	Overcast	14	Normal	Strong	Yes
9	Sunny	15	Normal	Weak	Yes
10	Rain	22	Normal	Weak	Yes
11	Sunny	19	Normal	Strong	Yes
12	Overcast	20	High	Strong	Yes
13	Overcast	33	Normal	Weak	Yes

	Outlook	Temp	Humidity	Wind	Play
1	Sunny	30	High	Weak	No
2	Sunny	31	High	Strong	No
6	Rain	13	Normal	Strong	No
8	Sunny	25	High	Weak	No
14	Rain	18	High	Strong	No



NB with Real Values

- Measurement values -> modeled with the Gaussian dist.
  - Assumption: Temperature of each class will follow the Gaussian distribution

	Outlook	Temp	Humidity	Wind	Play
3	Overcast	28	High	Weak	Yes
4	Rain	23	High	Weak	Yes
5	Rain	10	Normal	Weak	Yes
7	Overcast	14	Normal	Strong	Yes
9	Sunny	15	Normal	Weak	Yes
10	Rain	22	Normal	Weak	Yes
11	Sunny	19	Normal	Strong	Yes
12	Overcast	20	High	Strong	Yes
13	Overcast	33	Normal	Weak	Yes

$\mu = 20.4$ $\sigma = 7.1$

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$P(T = 22|yes) \propto \frac{1}{7.1 \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{22-20.4}{7.1}\right)^2} = 0.055$$

NB with Real Values

- Measurement values -> modeled with the Gaussian dist.
  - Assumption: Temperature of each class will follow the Gaussian distribution

	Outlook	Temp	Humidity	Wind	Play
1	Sunny	30	High	Weak	No
2	Sunny	31	High	Strong	No
6	Rain	13	Normal	Strong	No
8	Sunny	25	High	Weak	No
14	Rain	18	High	Strong	No

$\mu = 23.4 \quad \sigma = 7.8$

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$P(T = 22|no) \propto \frac{1}{7.8 \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{22-23.4}{7.8}\right)^2} = 0.050$$

NB with Real Values

From observation

- If there are continuous values, how can I ESTIMATE the probability?

$P(yes) = 9/14$

$P(no) = 5/14$

$P(sunny|yes) = 2/9$

$P(T = 22|yes) = 0.055$

$P(high|yes) = 3/9$

$P(strong|yes) = 3/9$

$P(sunny|no) = 3/5$

$P(T = 22|no) = 0.050$

$P(high|no) = 4/5$

$P(strong|no) = 3/5$

$$\frac{P(sunny|yes)P(T = 22|yes)P(high|yes)P(strong|yes)P(yes)}{P(sunny, mild, high, strong)} = \frac{1}{\alpha} \times \frac{2}{9} \times 0.055 \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}$$

$$\frac{P(sunny|no)P(T = 22|no)P(high|\text{no})P(strong|no)P(no)}{P(sunny, mild, high, strong)} = \frac{1}{\alpha} \times \frac{3}{5} \times 0.050 \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14}$$

NB with Real Values

- Naïve Bayesian Classifier with Counts
  - You observe wind blows each day
  - We count strong winds and weak winds in a day

from Count

one-event (하루하루 관찰) = Wind

	Outlook	Temp	Humidity	S-Wind	W-Wind	Play
1	Sunny	Hot	High	1	3	No
2	Sunny	Hot	High	4	2	No
3	Overcast	Hot	High	2	5	Yes
4	Rain	Mild	High	1	3	Yes
5	Rain	Cool	Normal	2	2	Yes
6	Rain	Cool	Normal	3	2	No
7	Overcast	Cool	Normal	5	1	Yes
8	Sunny	Mild	High	2	3	No
9	Sunny	Cool	Normal	1	2	Yes
10	Rain	Mild	Normal	0	4	Yes
11	Sunny	Mild	Normal	3	1	Yes
12	Overcast	Mild	High	3	2	Yes
13	Overcast	Hot	Normal	1	2	Yes
14	Rain	Mild	High	3	3	No

NB with Real Values

- Today is
  - O=Sunny, T=Mild, H=High, S-Wind=3, and W-Wind=2

- Will he play tennis today?

SW-WW
independent; 강한 상관관계가 있어서 하나로처리

$$P(Yes|O = Sunny, T = Mild, H = High, SW = 3, WW = 2)$$

$$= \frac{P(Sunny\,|\,yes)P(Mild\,|\,yes)P(High\,|\,yes)P(SW = 3, WW = 2|yes)P(yes)}{P(sunny, mild, high, 3,2)}$$

$$P(No|O = sunny, T = \text{Mild}, H = High, SW = 3, WW = 2)$$

$$= \frac{P(Sunny\,|\,no)P(Mild\,|\,no)P(High\,|\,\text{no})P(SW = 3, WW = 2|no)P(no)}{P(sunny, mild, high, 3,2)}$$

- Gaussian dist. is not good for modeling counts

NB with Real Values

- Naïve Bayesian Classifier with Counts

# of SW

$P(yes) = 9/14$

$P(no) = 5/14$

$P(sunny|yes) = 2/9$

$P(mild|yes) = 4/9$

$P(high|yes) = 3/9$

$P(SW = 3, WW = 2|yes) = ?$

$P(sunny|no) = 3/5$

$P(mild|no) = 2/5$

$P(high|no) = 4/5$

$P(SW = 3, WW = 2|no) = ?$

	Outlook	Temp	Humidity	S-Wind	W-Wind	Play
1	Sunny	Hot	High	1	3	No
2	Sunny	Hot	High	3	1	No
3	Overcast	Hot	High	2	5	Yes
4	Rain	Mild	High	1	3	Yes
5	Rain	Cool	Normal	1	3	Yes
6	Rain	Cool	Normal	3	1	No
7	Overcast	Cool	Normal	3	3	Yes
8	Sunny	Mild	High	2	2	No
9	Sunny	Cool	Normal	1	2	Yes
10	Rain	Mild	Normal	0	4	Yes
11	Sunny	Mild	Normal	3	1	Yes
12	Overcast	Mild	High	3	2	Yes
13	Overcast	Hot	Normal	1	2	Yes
14	Rain	Mild	High	3	1	No

NB with Real Values

- Naïve Bayesian Classifier with Counts

ML: 모든 구성요소는 simple
But, 그들이 complex하게 연결되어있다.

$P(yes) = 9/14$

$P(no) = 5/14$

$P(sunny|yes) = 2/9$

$P(mild|yes) = 4/9$

$P(high|yes) = 3/9$

$P(SW = 3, WW = 2|yes) = C(5,3) \times P(SW|yes)^3 \times P(WW|yes)^2$ → multinomial (동전 던지기 같은 것)

↳ ${}_5C_3$ = SW 3번이 들어갈 자리를 선택하기만하면됨

S S S WW

$P(sunny|no) = 3/5$

$P(mild|no) = 2/5$

$P(high|no) = 4/5$

$P(SW = 3, WW = 2|no) = C(5,3) \times P(SW|no)^3 \times P(WW|no)^2$

NB with Real Values

- We assume that Counts follows Multinomial dist.

	Outlook	Temp	Humidity	S-Wind	W-Wind	Play
3	Overcast	Hot	High	2	5	Yes
4	Rain	Mild	High	1	3	Yes
5	Rain	Cool	Normal	1	3	Yes
7	Overcast	Cool	Normal	3	3	Yes
9	Sunny	Cool	Normal	1	2	Yes
10	Rain	Mild	Normal	0	4	Yes
11	Sunny	Mild	Normal	3	1	Yes
12	Overcast	Mild	High	3	2	Yes
13	Overcast	Hot	Normal	1	2	Yes

- Estimate $P(SW|yes)$ and $P(WW|yes)$

$$P(SW|yes) = \frac{\#\ of\ Strong\ winds}{\#\ of\ All\ winds} = \frac{2+1+1+\cdots}{(2+5)+(1+3)+(1+3)+\cdots} = \frac{15}{40}$$

$$P(WW|yes) = \frac{\#\ of\ Weak\ winds}{\#\ of\ All\ winds} = \frac{5+3+3+\cdots}{(2+5)+(1+3)+(1+3)+\cdots} = \frac{25}{40}$$

NB with Real Values

- We assume that Counts follows Multinomial dist.

	Outlook	Temp	Humidity	S-Wind	W-Wind	Play
1	Sunny	Hot	High	1	3	No
2	Sunny	Hot	High	3	1	No
6	Rain	Cool	Normal	3	1	No
8	Sunny	Mild	High	2	2	No
14	Rain	Mild	High	3	1	No

  - Estimate $P(SW|no)$ and $P(WW|no)$

$$P(SW|no) = \frac{\#\ of\ Strong\ winds}{\#\ of\ All\ winds} = \frac{1+3+3+2+3}{(1+3)+(3+1)+(3+1)+\cdots} = \frac{12}{20}$$

$$P(WW|no) = \frac{\#\ of\ Weak\ winds}{\#\ of\ All\ winds} = \frac{3+1+1+2+1}{(1+3)+(3+1)+(3+1)+\cdots} = \frac{8}{20}$$

Summary

- Conditional independence assumption is often violated but it is easy to implement and works surprisingly well anyway

- When to use
  - Moderate or large training set available
  - Attributes that describe instances are conditionally independent given classification

- Successful applications
  - Diagnosis
  - Classifying text documents: spam or non-spam