

Naive Bayes

1 Naive Bayes

목표

1.

Naive Bayes

1. 확률의 개념
2. 베이즈 정리
3. Naive Bayes

확률

1.

Naive Bayes

1. 확률
2. 조건부 확률
3. 조건부 독립

확률

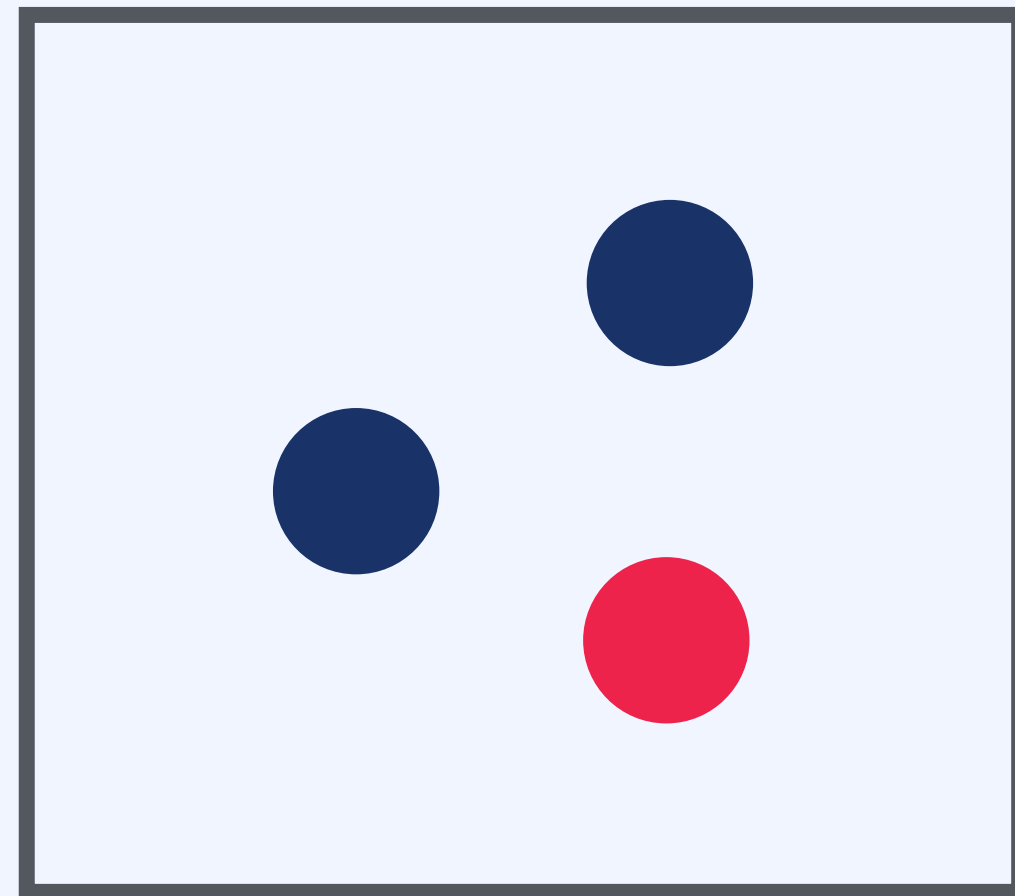
확률의 정의

1.

Naive Bayes

확률

특정한 사건이 일어날 가능성



파란 공이 뽑힐 확률: $2/3$

빨간 공이 뽑힐 확률: $1/3$

확률

조건부 확률

1.

Naive Bayes

조건부 확률

어떤 사건 A가 일어 났을 때, 다른 사건 B가 발생할 확률

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

조건부 확률 곱셈 공식

$$P(B|A)P(A) = P(B \cap A) = P(A \cap B) = P(A|B)P(B)$$

확률

독립

1.
Naive Bayes

독립

사건 A가 일어나는 것에 상관없이 사건 B가 일어날 확률이 일정할 때

$$P(A \cap B) = P(A)P(B)$$

조건부 독립

사건 C가 일어났을 때 서로 다른 사건 A, B가 독립일 때

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

베이지 정리

수식

1.
Naive Bayes

베이지 정리의 수식

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

Speed (X)	Result (Y)
Fast	Win
Fast	Win
Slow	Lose
Fast	Lose
Slow	Win
Fast	?

Speed (X)	Result (Y)
Fast	Win
Fast	Win
Slow	Lose
Fast	Lose
Slow	Win
Fast	?

$$P(Y = win) = \frac{3}{5}$$

$$P(Y = lose) = \frac{2}{5}$$

베이지 정리

예제

1.

Naive Bayes

Speed (X)	Result (Y)
Fast	Win
Fast	Win
Slow	Lose
Fast	Lose
Slow	Win
Fast	?

$$P(X = fast) = \frac{3}{5}$$

$$P(X = slow) = \frac{2}{5}$$

베이지 정리

예제

1.

Naive Bayes

Speed (X)	Result (Y)
Fast	Win
Fast	Win
Slow	Lose
Fast	Lose
Slow	Win
Fast	?

$$P(X = fast|Y = win) = \frac{2}{3}$$

$$P(X = slow|Y = win) = \frac{1}{3}$$

Speed (X)	Result (Y)
Fast	Win
Fast	Win
Slow	Lose
Fast	Lose
Slow	Win
Fast	?

$$P(X = fast|Y = lose) = \frac{1}{2}$$

$$P(X = slow|Y = lose) = \frac{1}{2}$$

베이지 정리

예제

1.

Naive Bayes

Speed (X)	Result (Y)
Fast	Win
Fast	Win
Slow	Lose
Fast	Lose
Slow	Win
Fast	?

$$P(Y = win|X = fast) = \frac{P(X = fast|Y = win)P(Y = win)}{P(X = fast)}$$

$$P(Y = lose|X = fast) = \frac{P(X = fast|Y = lose)P(Y = lose)}{P(X = fast)}$$

베이지 정리

예제

1.

Naive Bayes

Speed (X)	Result (Y)
Fast	Win
Fast	Win
Slow	Lose
Fast	Lose
Slow	Win
Fast	?

$$P(Y = win|X = fast) = \frac{P(X = fast|Y = win)P(Y = win)}{P(X = fast)}$$

$$= \frac{2/3 \times 3/5}{3/5} = \frac{2}{3}$$

$$P(Y = lose|X = fast) = \frac{P(X = fast|Y = lose)P(Y = lose)}{P(X = fast)}$$

$$= \frac{1/2 \times 2/5}{3/5} = \frac{1}{3}$$

베이지 정리

예제

1.

Naive Bayes

Speed (X)	Result (Y)
Fast	Win
Fast	Win
Slow	Lose
Fast	Lose
Slow	Win
Fast	Win!

$$P(Y = win|X = fast) = \frac{P(X = fast|Y = win)P(Y = win)}{P(X = fast)}$$

$$= \frac{2/3 \times 3/5}{3/5} = \frac{2}{3}$$

$$P(Y = lose|X = fast) = \frac{P(X = fast|Y = lose)P(Y = lose)}{P(X = fast)}$$

$$= \frac{1/2 \times 2/5}{3/5} = \frac{1}{3}$$

베이지 정리

한계점

1.

Naive Bayes

Speed (X1)	Condition(X2)	Weather(X3)	Result (Y)
Fast	Good	Hot	Win
Fast	Good	Hot	Win
Slow	Bad	Hot	Lose
Fast	Bad	Cold	Lose
Slow	Good	Cold	Win
Fast	Good	Hot	?

$$P(X = x|Y = y) = P(X1 = fast, X2 = good, X3 = hot|Y = win)$$

$$P(X = x|Y = y) \text{ for all } x, y \rightarrow (2^d - 1) \quad k, d: \text{관측치 수}, k: \text{class 수}$$

계산량이 너무 많다!

계산량이 너무 많다!

조건부 독립을 가정

$$P(X = x|Y = y) = P(X1 = fast, X2 = good, X3 = hot|Y = win)$$

$$P(X = x|Y = y) = P(X1 = fast|Y = win)P(X2 = good|Y = win)P(X3 = hot|Y = win)$$

Naive Bayes

정의

1.

Naive Bayes

Naive Bayes

- 종속 변수(Y)가 주어졌을 때 입력 변수(X)들은 모두 조건부 독립이다.
- 예측 변수들의 정확한 조건부 확률은 각 조건부 확률의 곱으로 충분히 잘 추정 할 수 있다는 단순한 가정
- 데이터셋을 순진하게 믿는다! -> Naïve(순진한) Bayes!

Naive Bayes

예시

1.
Naive Bayes

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	1	0
1	0	1	0
1	0	0	1
0	0	1	?

Naive Bayes

예시

1. Naive Bayes

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	1	0
1	0	1	0
1	0	0	1
0	0	1	?



$$P(Y = 1 | X1 = 0, X2 = 0, X3 = 1)$$

$$= P(X1 = 0 | Y = 1)P(X2 = 0 | Y = 1)P(X3 = 1 | Y = 1)P(Y = 1)$$

Naive Bayes

예시

1.
Naive Bayes

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	1	0
1	0	1	0
1	0	0	1
0	0	1	?

$P(Y = 1|X1 = 0,X2 = 0,X3 = 1)$
 $= P(X1 = 0|Y = 1)P(X2 = 0|Y = 1)P(X3 = 1|Y = 1)P(Y = 1)$

Naive Bayes

예시

1.
Naive Bayes

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	1	0
1	0	1	0
1	0	0	1
0	0	1	?

$P(Y = 1|X1 = 0,X2 = 0,X3 = 1)$
 $= P(X1 = 0|Y = 1)P(X2 = 0|Y = 1)P(X3 = 1|Y = 1)P(Y = 1)$
 $P(X1 = 0|Y = 1) = \frac{2}{4}$

Naive Bayes

예시

1.
Naive Bayes

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	1	0
1	0	1	0
1	0	0	1
0	0	1	?

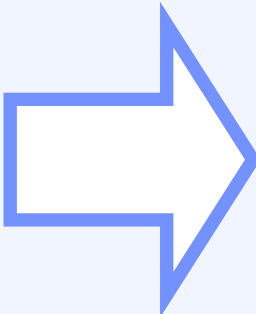
$P(Y = 1|X1 = 0,X2 = 0,X3 = 1)$
 $= P(X1 = 0|Y = 1)P(X2 = 0|Y = 1)P(X3 = 1|Y = 1)P(Y = 1)$

$P(X1 = 0|Y = 1) = \frac{2}{4}$

$P(X2 = 0|Y = 1) = \frac{2}{4}$

$P(X3 = 1|Y = 1) = \frac{1}{4}$

$P(Y = 1) = \frac{4}{7}$

 $\frac{2}{4} * \frac{2}{4} * \frac{1}{4} * \frac{4}{7} = \frac{1}{28} = 0.036$

Naive Bayes

예시

1.
Naive Bayes

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	1	0
1	0	1	0
1	0	0	1
0	0	1	?

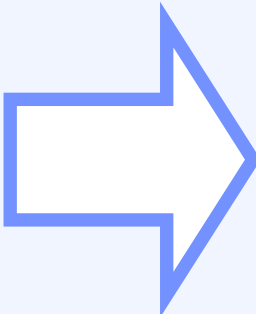
$P(Y = 0|X1 = 0,X2 = 0,X3 = 1)$
 $= P(X1 = 0|Y = 0)P(X2 = 0|Y = 0)P(X3 = 1|Y = 0)P(Y = 0)$

$P(X1 = 0|Y = 0) = \frac{1}{3}$

$P(X2 = 0|Y = 0) = \frac{2}{3}$

$P(X3 = 1|Y = 0) = 1$

$P(Y = 0) = \frac{3}{7}$

 $\frac{1}{3} * \frac{1}{3} * 1 * \frac{3}{7} = \frac{2}{21} = 0.095$

Naive Bayes

예시

1.
Naive Bayes

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	1	0
1	0	1	0
1	0	0	1
0	0	1	0

$P(Y = 1|X1 = 0,X2 = 0,X3 = 1) = 0.036$

$P(Y = 0|X1 = 0,X2 = 0,X3 = 1) = 0.095$

Laplace Smoothing

Laplace Smoothing

- Count하다 보면 한 번도 나오지 않는 경우도 있을 수 있다.
-> 확률이 0이 되는 것을 방지해야 함
- 최소한의 확률을 정해 준다.

$$P(x|c) = \frac{\text{count}(x,c)+1}{\sum_{x \in v} \text{count}(x,c)+v}, \quad v = \text{입력변수의 개수}$$

Naive Bayes

장단점

1.

Naive Bayes

장점

- 변수가 많을 때 좋다.
- 텍스트 데이터에서 큰 강점을 보인다.

- 단점

- 희귀한 확률이 나왔을 때 처리하기 힘들다.
- 조건부 독립이라는 가정 자체가 비현실적