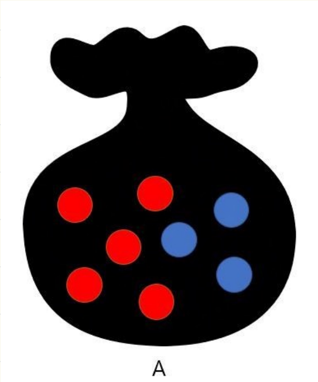
확률

특정 사건이 일어날 가능성



빨간공 뽑을 확률 : $\frac{5}{8}$

파란공 뽑을 확률 : $\frac{3}{8}$

조건부 확률

어떤 사건이 일어난 조건에서, 다른 사건이 일어날 확률

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

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

독립과 조건부 독립

독립: 한 사건이 일어날 확률이 다른 사건이 일어날 확률에 영향을 미치지 않는 상태

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

조건부독립: 한 사건이 일어났다는 가정하에서, 서로 다른 두 사건은 독립인 상황

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

베이즈 정리

두 확률 변수의 사전확률과 사후확률 사이의 관계를 나타내는 정리

사전 확률로부터 사후확률을 구하고자 한다.

$$\underbrace{P(H|D)}_{\text{사후확률}} = \frac{\overbrace{P(D|H)}^{\text{우도}}\,\overbrace{P(H)}^{\text{사전확률}}}{\underbrace{P(D)}_{\text{정규화 상수}}}$$

예제

Sky (X)	Enjoy Point (Y)
Sunny	Yes
Sunny	Yes
Rainy	No
Sunny	No
Rainy	Yes

③ $$P(Y=yes|X=sunny) = \frac{\overset{②}{P(X=sunny|Y=yes)}\,\overset{①}{P(Y=yes)}}{P(X=sunny)}$$

굳이 구할필요X

① prior

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

② Likelihood

$$P(X=sunny|Y=yes) = \frac{P(X=sunny\cap Y=yes)}{P(Y=yes)} = \frac{\frac{2}{5}}{\frac{3}{5}} = \frac{2}{3}$$

③ Posterior

$$P(Y=yes|X=sunny) = ① \times ② = \frac{3}{5}\times\frac{2}{3} = \frac{2}{5}$$

문제점

X 변수가 많을수록 계산량이 많아진다.

$P(X=x|Y=y)$ for all $x, y \rightarrow (2^d-1)^k$

해결책 : 조건부 독립을 가정

Naive Bayes Classfication

가정 : 종속변수 Y가 주어졌을 때, 입력 변수들이 모두 독립이다.

$$f^*(x) = \text{argmax}_{Y=y}\, P(X=x|Y=y)\,P(Y=y)$$

$$\approx \text{argmax}_{Y=y}\, P(Y=y)\prod_{1\le i\le d} P(X_i=x_i|Y=y)$$

장점

- 입력 공간의 차원이 높을 때 유리
- 텍스트에 강점
- 가우시안 나이브베이즈를 활용하면 input이 연속형일 때도 사용가능

단점

- 희귀한 확률이 나왔을 때 (라플라스 스무딩)
- 조건부 독립이라는 가정 자체가 비현실적