

- 남준연 -

< Naive Bayes classifier >

- 확률 : 특정한 사건이 일어날 가능성
 - 조건부 확률 : 어떤 사건이 일어난 조건 하에서, 다른 사건이 일어날 확률
- $$P(B|A) = \frac{P(A \cap B)}{P(A)} \Rightarrow P(B|A)P(A) = P(B \cap A) = P(A \cap B) = P(A|B)P(B)$$

- 독립 : $P(A \cap B) = P(A)P(B)$
- 조건부 독립 : $P(A, B|C) = P(A|C)P(B|C)$

베이지 정리 : 두 확률변수의 사전 확률과 사후확률 사이의 관계를 나타내는 정리

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

↓
사후확률

→ 사전확률

→ 사건 D의 발생가능성

→ $P(D|H)$: Likelihood.
사전확률의 라게 경향을 잘 설명하는 정도

문제점 : 계산량이 많아짐 \Rightarrow 해결책 : 조건부 독립을 가정

< Naive Bayes classification >

가정 : 종속변수(y)가 주어졌을 때, 입력 변수들이 모두 독립이다!

$$f^*(x) = \underset{y=y}{\operatorname{argmax}} P(x=x | y=y) P(y=y) \approx \underset{y=y}{\operatorname{argmax}} P(y=y) \prod_{i=1}^d P(x=x_i | y=y)$$

장점 ① 알아야 할 파라미터의 수가 대폭 줄다

② Feature들의 종류 바뀌면서 계산이 수월해진다

단점 ① 희귀한 확률 (라플라스 스무딩) 나눌 때

② 조건부 독립이라는 가정이 비현실적