Naïve Bayes Classifier

1. 확률 기초

- 조건부확률 : 어떤 사건이 일어난 조건 하에서, 다른 사건이 일어날 확률 $P(B|A) = \dfrac{P(B \cap A)}{P(A)}$
- 독립 : 한 사건이 일어날 확률이 다른 사건이 일어날 확률에 영향을 미치지 않는 상태
  - ↳ $P(A \cap B) = P(A) \times P(B)$
- 조건부 독립 : 한 사건이(c) 일어났다는 가정하에서, 서로 다른 두 사건은 독립인 상황
  - ↳ $P(A, B | C) = P(A|C)\,P(B|C)$

2. 베이즈 정리

↳ 두 확률 변수의 사전확률과 사후 확률 사이의 관계를 나타내는 정리

⇒ 사전 확률로부터 사후확률을 구하고자 한다.

$$\underbrace{P(H|D)}_{\text{posterior}} = \frac{\underbrace{P(D|H)}_{\text{Likelihood}} \times \underbrace{P(H)}_{\text{Prior}}}{P(D)}$$

Likelihood : 사전확률의 '과거 경험'을 잘 설명하는 정도

But .. class가 많아질 수록 계산이 복잡해짐!

⇓

조건부 독립 가정!

3. 조건부 독립

: 가정) 종속 변수(Y)가 주어졌을 때, 입력 변수들이 모두 독립이다.

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

↳ 계산이 대폭 쉬워짐!

4. 라플라스 스무딩

: likelihood가 0이 되는 것을 방지하도록 최소한의 확률 정해주자!

① P_{LAP} : 실제보다 한 번씩 더 관찰되었다고 가정하기

$$P_{LAP} = \frac{C(x) + 1}{\sum_x [C(x) + 1]}$$

② $P(x|c)$: 분자에 1 더하기, 분모엔 입력 변수들의 개수 V 더해주기

$$P(x|c) = \frac{\text{count}(x, c) + 1}{\sum \text{count}(x, c) + V}$$