[베이즈 정리]

두 확률 변수의 사전 확률로부터 사후 확률을 구하는 정리

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

-> Posterior = Prior*Likelihood/Normalizing Constant

-> 사전에 예상한 가설 H에 대한 확률(사전 확률)에 관찰된 데이터 D로 계산한 가능성 정도(Likelihood)를 곱하고 관측결과가 발생할 확률(Normalizing Constant)로 나눠서 구함

[나이브 베이즈]

베이즈 정리 - 변수가 늘어날수록 연산량 증가

-> 나이브 베이즈로 해결 가능

- 가정: 종속변수가 주어졌을 때 입력 변수들이 모두 독립
- 결과가 주어졌을 때, 조건부 확률은 각 조건부 확률의 곱으로 추정 가능
- 장점
    - ■ 알아야할 파라미터 수 대폭 감소
    - ■ Feature들의 곱으로 바뀌면서 계산 수월
    - ■ 입력 공간의 차원이 높을 때 유리
    - ■ 텍스트에서 강점
    - ■ 가우시안 나이브베이즈를 활용하면 input이 연속형일 때도 사용 가능
- 단점
- ■ 희귀한 확률이 나왔을 때
    - ✓ 라플라스 스무딩 사용
        - ● Likelihood가 0이 되는 것을 방지하기 위함
        - ● 실제보다 한 번씩 더 관찰되었다고 가정하기
        - ● 분자에 1 더하고 분모에 입력변수들의 개수 v 더함
        - ● $P(x|c) = \frac{count(x,c)+1}{\sum_x^v count(x,c)+v}$
- ■ 조건부 독립이라는 가정 자체가 비현실적

EX. $P(H|D_1, D_2) = P(D_1, D_2|H) * P(H)/P(D_1, D_2)$

➔ $P(D_1, D_2|H) = P(D_1|H) * P(D_2|H)$

➔ $P(D_1, D_2) = P(D_1) * P(D_2)$