

Naive Bayes

- 1-1) 확률: 특정한 사건이 일어날 가능성을 나타낸 것
- 1-2) 조건부확률: 어떠한 사건이 일어났다는 조건 아래, 다른 사건이 일어날 확률
- 1-3) 독립: 한 사건이 일어날 확률이 다른 사건이 일어날 확률에 영향을 미치지 않는 상태
조건부 독립: 한 사건이 일어났다는 가정 아래, 서로 다른 두 사건이 독립인 상황
- 2-1) 베이즈 정리: 두 확률변수의 사전확률과 사후확률 사이의 관계를 나타내는 정리
 - 사전확률(Prior): 과거의 경험을 토대로 추정한 H의 확률 / $P(H)$
 - 사후확률(Posterior): 관측결과 알고 있는 정보가 일어났을 때의 H의 확률 / $P(H|D)$
 - 우도(Likelihood): 사전확률의 과거 경험을 잘 설명하는 정도 / $P(D|H)$
H를 바탕으로 하는 관측결과 사건의 확률
- 3-1) 변수가 늘어날수록, 기하급수적으로 연산량이 증가하는데, 차원을 줄이고 싶지 않을 때, 조건부 독립을 가정
- 3-2) Naïve Bayes Classification
 - 가정: 종속변수가 주어졌을 때, 입력 변수들이 모두 독립이라는 조건부 독립을 가정
 - 결과가 주어졌을 때, 예측 변수 벡터의 정확한 조건부 확률은 각 조건부 확률의 곱으로 충분히 추정할 수 있다는 단순한 가정을 기초로 함
 - 알아야할 파라미터의 수가 많이 줄어들게 된다: $(2^d - 1) \rightarrow d$
 - Feature들의 곱으로 바뀌면서 계산이 수월해짐
- 3-3) 라플라스 스무딩
likelihood가 0이 되는 것을 방지하도록 최소한의 확률을 정해주는 것

나이브 베이즈의 장단점

장점

입력 공간의 차원이 높을 때, 알아야할 파라미터의 수를 줄일 수 있어 유리

텍스트에서 강점을 보임

Gaussian Naïve Bayes를 사용한다면, 연속형 변수에서도 사용가능

단점

조건부 독립이라는 가정 자체가 비현실적

