나이브 베이즈 스팸필터

부산대학교 정보·의생명공학대학
정보컴퓨터공학부

개요

- ❖ 베이즈 정리
- ❖ 나이브 베이즈를 이용한 스팸필터
- ❖ 나이브 베이즈를 이용한 스팸필터 예제 코드

베이즈 정리

확률의 의미

❖ 빈도주의(Frequenctist) 관점

- 반복적으로 선택된 표본이 사건(부분 집합) A의 원소가 될 경향(propensity)
- 동전을 반복하여 던졌을 경우 전체 횟수에 확률값을 곱한 숫자만큼 해당 사건이 발생한다고 봄
- 10,000번 던지면 10,000 x 0.5 = 5,000 번 앞면이 나오는 경향을 가진다는 의미

❖ 베이지안(Bayesian) 관점

- 선택된 표본이 특정한 사건(부분 집합)에 속한다는
  가설(hypothesis), 명제(proposition) 혹은 주장(assertion)의 신뢰도(degree of belief)
- 반복이라는 개념이 사용되지 않음
- E.g., " 새는 날 수 있다." → 베이지안 관점: 새가 날 수 있는 가능성은 95%이다.
- E.g., 동전 던지기 → 베이지안 관전: '앞면이 나왔다'는 주장의 신뢰도가 0.5 이다.
- E.g., 4지 선다형 객관식 문제 풀 때: 표본집합→{1, 2, 3, 4}, 할당된 확륙→{0, 0.4, 0.4, 0.2}
  - 반복 또는 빈도와는 전혀 관계 없음, 확률 P({1})은 '정답이 1이다'라는 주장에 대한 신뢰도 일 뿐

베이즈 정리 (Bayes's Theorem)

❖ For any events A and B such that $\mathrm{P}(B) \neq 0$,

$$\boldsymbol{P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}} \quad \Rightarrow \quad \boldsymbol{P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}}$$

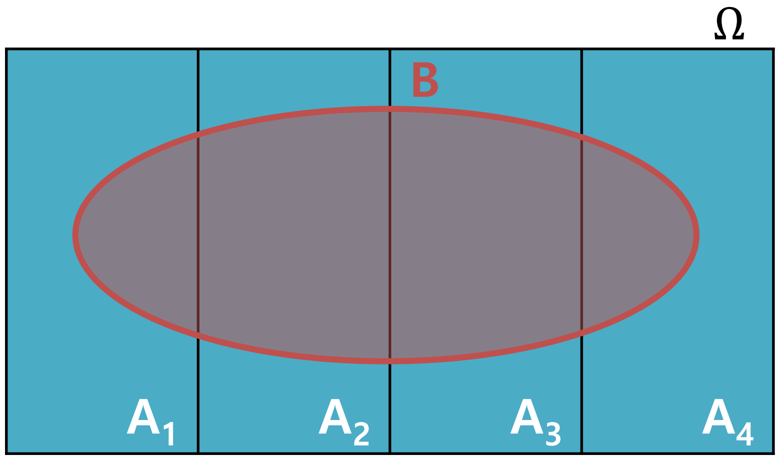
- $\boldsymbol{P(A|B)}$: 사후확률(posterior). 사건 B가 발생한 후 갱신된 사건 A의 확률
- $\boldsymbol{P(A)}$: 사전확률(prior). 사건 B가 발생하기 전에 가지고 있던 사건 A의 확률
- $\boldsymbol{P(B|A)}$: 가능도 또는 우도(Likelihood). 사건 A가 발생한 경우 사건 B의 확률
- $\boldsymbol{P(B)}$: 정규화 상수 또는 증거(Evidence). 확률의 크기 조정

❖ A method to update belief (of H) on the basis of new information (E)

베이즈 정리 확장 1

❖ 만약 사건 A_i 가 서로 배타적이고 완전하다면

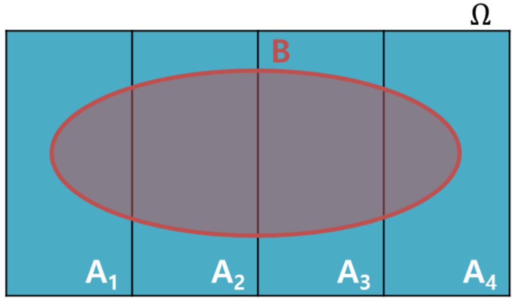
- $A_i \cap A_j = \emptyset$
- $A_1 \cup A_2 \cup \cdots = \Omega$



❖ 전체 확률의 법칙 (law of total probability) 이용하여 베이즈 정리 확장

- $$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B)} = \frac{P(B|A_1)P(A_1)}{\sum_i P(A_i, B)} = \frac{P(B|A_1)P(A_1)}{\sum_i P(B|A_i)P(A_i)}$$

베이즈 정리 확장 1 (cont.)



- $$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) + P(B|A_4)P(A_4)}$$

  $$P(A_2|B) = \frac{P(B|A_2)P(A_2)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) + P(B|A_4)P(A_4)}$$

  $$\vdots$$

- $P(A_1|B) \propto P(B|A_1)P(A_1)$

  $$\vdots$$

- $A_1 = A,\ A_2 = A^C$ 인 경우 (클래스가 2개인 경우)

  $$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\ &= \frac{P(B|A)P(A)}{P(B,A) + P(B,A^C)} \\ &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)} \\ &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)(1 - P(A))} \end{aligned}$$

예제 1: 검사 시약 문제

❖ 제약사에서 환자가 특정한 병에 걸린 지 확인하는 시약을 만들었다. 그 병에 걸린 환자에게 시약을 테스트한 결과 99%의 확률로 양성 반응을 보였다. 병에 걸린 지 확인이 되지 않은 어떤 환자가 이 시약을 테스트한 결과 양성 반응을 보였다면 이 환자가 그 병에 걸려 있을 확률은 얼마인가? 99%일까?

- 사건
  - 병에 걸리는 경우: 사건 D
  - 양성 반응을 보이는 경우: 사건 S
  - 병에 걸린 사람이 양성 반응을 보이는 경우: 조건부 사건 S|D
  - 양성 반응을 보이는 사람이 병에 걸려 있을 경우: 조건부 사건 D|S
- 문제
  - P(S|D) = 0.99 가 주어졌을 때, P(D|S) = ?
- 입수된 추가 정보
  - 전체 인구 중 이 병에 걸린 사람은 0.2% → P(D) = 0.002
  - 병에 걸리지 않은 사람에게 시약 검사했을 경우 False Positive 나타날 확률 5% → $P(S|D^C) = 0.05$

$$
\begin{aligned}
P(D|S) &= \frac{P(S|D)P(D)}{P(S)} \\
&= \frac{P(S|D)P(D)}{P(S,D) + P(S,D^C)} \\
&= \frac{P(S|D)P(D)}{P(S|D)P(D) + P(S|D^C)P(D^C)} \\
&= \frac{P(S|D)P(D)}{P(S|D)P(D) + P(S|D^C)(1 - P(D))} \\
&= \frac{0.99 \cdot 0.002}{0.99 \cdot 0.002 + 0.05 \cdot (1 - 0.002)} \\
&= 0.038
\end{aligned}
$$

베이즈 정리 확장 2

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

- 베이즈 정리는 사건 A의 확률이 사건 B에 의해 갱신(update)된 확률을 계산함
- 이때 추가적인 사건 C가 발생했다면,

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

- $P(A|B,C)$ 는 B와 C가 조건인 A의 확률, 즉, $P(A|(B \cap C))$
- 증명

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

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

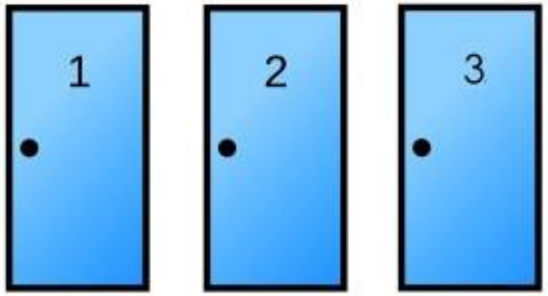
$$P(A|B,C)P(C|B)P(B) = P(C|A,B)P(A|B)P(B)$$

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

몬티 홀 문제 (Monty Hall problem)

- 세 개의 문 중에 하나를 선택하여 문 뒤에 있는 선물을 가지는 게임쇼에 참가했다. 한 문 뒤에는 자동차가 있고, 나머지 두 문 뒤에는 염소가 있다. 이때 어떤 사람이 예를 들어 1번 문을 선택했을 때, 게임쇼 진행자는 3번 문을 열어 문뒤에 염소가 있음을 보여주면서 1번 대신 2번을 선택하겠냐고 물었다.

  참가자가 자동차를 가지려할 때 원래 선택했던 번호를 바꾸는 것이 유리할까?

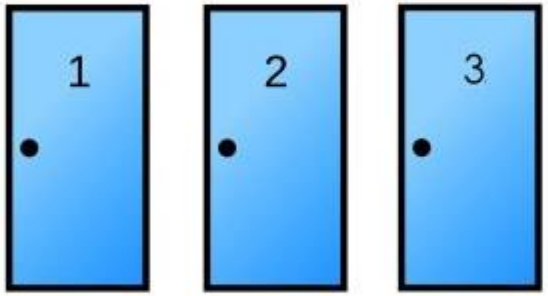


- C: 자동차가 있는 문을 나타내는 확률변수. 표본공간: {0, 1, 2}
- X: 참가자가 선택한 문을 나타내는 확률변수. 표본공간: {0, 1, 2}
- H: 진행자가 열어준 문을 나타내는 확률변수. 표본공간: {0, 1, 2}

- 만약, 참가자가 1번문 선택, 진행자가 2번문 열어서 염소 보여준경우 → X_1, H_2
  - 이때 자동차는 0 아니면 1번 문 뒤에 있는 이진 분류 문제

- 두가지 사실
  - P(C,X) = P(C)P(X) → 즉, 자동차 위치와 참가자 선택은 서로 독립적
  - 진행자의 선택은 참가자의 선택에 종속적

몬티 홀 문제 (Monty Hall problem) (cont.)



- 참가자 1번문 선택, 진행자 2번 문 열어 염소확인 인 경우
  0번 문 뒤에 차가 있을 확률은

$$
\begin{aligned}
P(C_0 \mid X_1, H_2) &= \frac{P(C_0, X_1, H_2)}{P(X_1, H_2)} \\
&= \frac{P(H_2 | C_0, X_1) P(C_0, X_1)}{P(X_1, H_2)} \\
&= \frac{P(C_0) P(X_1)}{P(H_2 | X_1) P(X_1)} \quad = \frac{P(C_0)}{P(H_2 | X_1)} \\
&= \frac{P(C_0)}{P(H_2, C_0 | X_1) + P(H_2, C_1 | X_1) + P(H_2, C_2 | X_1)} \\
&= \frac{P(C_0)}{P(H_2 | X_1, C_0) P(C_0) + P(H_2 | X_1, C_1) P(C_1) + P(H_2 | X_1, C_2) P(C_2)} \\
&= \frac{\frac{1}{3}}{1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} \\
&= \frac{2}{3}
\end{aligned}
$$