[S3-03] 4-1 베이지안 기법

베이즈 정리(Bayes' theorem)

- 두 확률 변수(A, B)의 사전 확률과 사후 확률 사이의 관계를 나타내는 정리
  - 사전 확률, P(A) : 특정 사건(B)이 일어나기 전의 확률, 관측자가 관측 전 가지고 있는 확률 분포
  - 사후 확률, P(A|B) : 사전 확률 P(A)과 가능도 P(B|A)가 주어지면, 관측 값을 얻은 다음 베이즈 정리에 의해 얻을 수 있는 확률
- A가 불확실성을 계산해야 하는 대상, B가 관측하여 알아낼 수 있는 대상일 때
- A의 확률은 B가 관측된 후 사전확률 P(A)에서 사후확률 P(A|B)로 변화함
- 베이즈 정리 기반의 지도 학습 분류 모델로 나이브 베이즈(Naïve Bayes) 모델이 있음

A가 주어졌을 때, B의 조건부 확률 (Likelihood, 가능도)

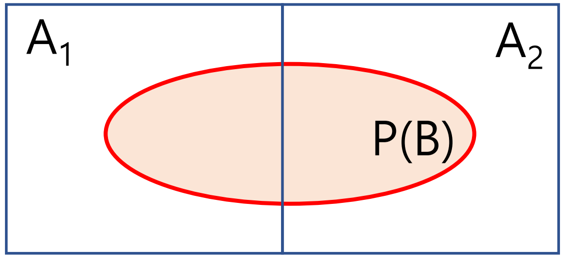
A의 사전 확률 (Prior)

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

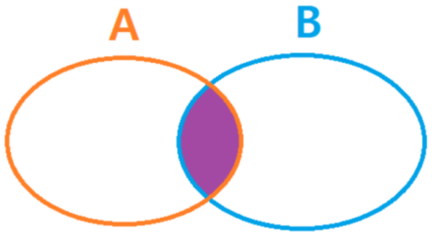
B가 주어진 경우 A의 사후 확률 (Posterior)

B의 사전 확률 (Evidence, 증거)

$$\mathrm{P}(\boldsymbol{A_1}|\boldsymbol{B}) = \frac{\mathrm{P}(\boldsymbol{B}|\boldsymbol{A_1})\mathrm{P}(\boldsymbol{A_1})}{\mathrm{P}(\boldsymbol{B}|\boldsymbol{A_1})\ \mathrm{P}(\boldsymbol{A_1}) + \mathrm{P}(\boldsymbol{B}|\boldsymbol{A_2})\mathrm{P}(\boldsymbol{A_2}) + \ldots}$$



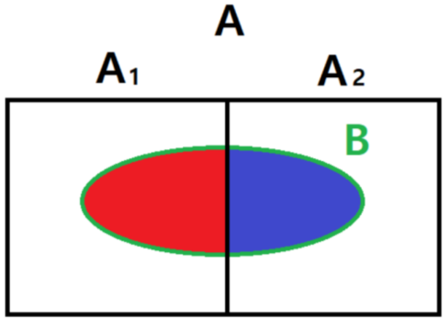
조건부 확률



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

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

전체 확률의 법칙



사건 A_i 가 서로 배타적이고 완전할 때

$A_i \cap A_j \ldots = 0$: 교집합이 없음

$A_i \cup A_j \ldots = \Omega$: 합집합이 표본공간

$$P(A_1|B) = \frac{P(B \cap A_1)}{P(B)} = \frac{P(B|A_1)P(A_1)}{P(B)}$$

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

곱셈 공식

$$P(B) = P(B \cap A) = P(B \cap A_1) + P(B \cap A_2)$$
$$= P(B|A_1)\ P(A_1) + P(B|A_2)P(A_2)$$

A가 n 개 일 경우

$$\frac{P(B|A_1)P(A_1)}{\sum_{i=1}^{n} P(B|A_i)\ P(A_i)}$$

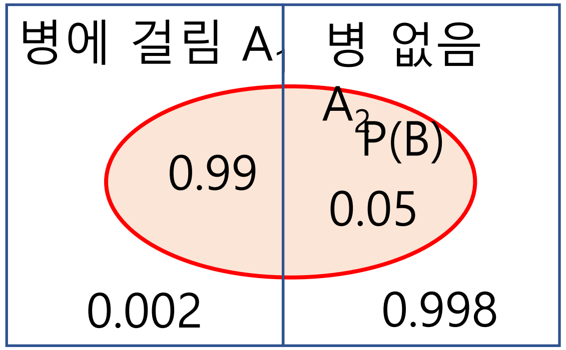
	A_1	A_2	A_3
P(B)	0.01	0.02	0.03
	0.5	0.3	0.2

[S3-03] 4-3 베이즈 정리

제약사에서 환자가 특정 병에 걸린 상태인지 확인하는 시약을 만들었다. 그 병에 걸린 환자에게 시약을 테스트한 결과 99%의 확률로 양성 반응을 보였다. 병에 걸린 것인지 확인이 되지 않은 어떤 환자가 이 시약을 테스트 한 결과 양성 반응을 보였다면, 이 환자가 그 병에 걸려 있을 확률은 얼마인가?

단, 이 병은 전체 인구 중 걸린 사람이 0.2%인 희귀병이며, 이 병에 걸리지 않은 사람에게 시약 검사를 했을 때 양성 반응이 나타날 확률이 5%이다

P(양성|병에 걸림)	병에 걸린 환자일 때 양성 반응일 확률 : 0.99
P(병에 걸림|양성)	양성 반응일 때 병에 걸린 환자일 확률 : ?



$$P(\boldsymbol{A}|\boldsymbol{B}) = \frac{P(\boldsymbol{B}|\boldsymbol{A})\, P(\boldsymbol{A})}{P(\boldsymbol{B})} = \frac{P(\boldsymbol{B}|\boldsymbol{A_1})P(\boldsymbol{A_1})}{P(\boldsymbol{B}|\boldsymbol{A_1})\, P(\boldsymbol{A_1}) + P(\boldsymbol{B}|\boldsymbol{A_2})P(\boldsymbol{A_2})}$$

$$P(\boldsymbol{A_1}|\boldsymbol{B}) = \frac{\mathbf{0.99} * \mathbf{0.002}}{\mathbf{0.99} * \mathbf{0.002} + \mathbf{0.05} * \mathbf{0.998}} = \mathbf{0.038}$$

[S3-03] 4-4 베이즈 정리 사용

- 멀티 클래스 분류에서 베이즈 정리 사용
  - 여러 배타적이고 완전한 사건 중 가장 확률이 높은 하나의 사건을 골라야 한다면
    4개의 A 중 B에 대한 조건부 확률이 가장 높은 사건을 고르는 것
  - 조건부 확률이 가장 높은 사건을 고르는 것이 목적이라면, 분모가 동일하므로 분자만 비교해도 됨

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

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

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

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

1. 어떤 공장에서 생산하는 제품의 종류가 A, B, C일 때 각각의 생산량은 50%, 30%, 20%이다. 제품 A의 불량률은 1%, 제품 B의 불량률은 2%, 제품 C의 불량률은 3%이다. 품질 검사 과정에서 불량품이 나온 경우, 해당 제품이 제품 A일 확률은?

① 29.41%
② 35.29%
③ 64.71%
④ 70.59%

A가 불량일 확률 / (A가 불량일 확률 + B가 불량일 확률 + C가 불량일 확률)
= (0.01* 0.5) / (0.01*0.5 + 0.02*0.3 + 0.03*0.2) = 0.2941

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

A_1	A_2	A_3
		P(B)
0.01	0.02	0.03
0.5	0.3	0.2

2. 새로운 질병 A가 발견되었다. 질병 A를 전체 국민의 P(A) 정도가 앓고 있으며, 전체 국민 중 어느 한 사람이 질병 A를 앓지 않는 건강한 사람일 확률은 P(B)라고 한다. 검진을 했을 때 질병 A에 걸린 사람을 정확히 검진할 확률이 P(x|A) 이며, 질병 A를 앓지 않는 건강한 사람이 질병 A에 걸렸다고 검진할 확률이 P(x|B)이다. P(A), P(B), P(x|A), P(x|B) 를 가지고, 어떤 사람을 검진해서 질병 A에 걸렸다는 결과가 나왔을 때, 그 사람이 실제로 질병 A를 앓지 않는 건강한 사람일 확률 P(B|x)의 수식을 구하시오.

1. $\dfrac{\mathrm{P(B)}}{\mathrm{P(x|B)P(B)}}$

2. $\dfrac{\mathrm{P(x|B)P(B)}}{\mathrm{P(A)} + \mathrm{P(B)}}$

3. $\dfrac{\mathrm{P(x|B)}}{\mathrm{P(x|A)P(x|B)}}$

4. $\dfrac{\mathrm{P(x|B)P(B)}}{\mathrm{P(x|A)\,P(A)} + \mathrm{P(x|B)P(B)}}$

베이즈 정리 공식

$$\mathrm{P}(\boldsymbol{A}|\boldsymbol{B}) = \frac{\mathrm{P}(\boldsymbol{B}|\boldsymbol{A})\,\mathrm{P}(\boldsymbol{A})}{\mathrm{P}(\boldsymbol{B})} = \frac{\mathrm{P}(\boldsymbol{B}|\boldsymbol{A}_1)\mathrm{P}(\boldsymbol{A}_1)}{\mathrm{P}(\boldsymbol{B}|\boldsymbol{A}_1)\,\mathrm{P}(\boldsymbol{A}_1) + \mathrm{P}(\boldsymbol{B}|\boldsymbol{A}_2)\mathrm{P}(\boldsymbol{A}_2) + \ldots}$$