

용어 정리

또, 식 (1)에 있는 공식에서 E와 H가 무엇인지 알아보자.

식 (1)에서 H는 Hypothesis의 약자로써 가설 혹은 '어떤 사건이 발생했다는 주장'을 의미한다.

거기에, 식 (1)에서 E는 Evidence의 약자로 '새로운 정보'를 의미한다.

따라서 $P(H)$ 는 어떤 사건이 발생했다는 주장에 관한 신뢰도, $P(H|E)$ 는 새로운 정보를 받은 후 갱신된 신뢰도를 의미한다.

(이의 존재하는 사건)

그리고 $P(H)$ 와 $P(H|E)$ 는 각각 사전 확률, 사후 확률이라고 부르는데, 사전(事前), 사후(事後)라는 단어를 생각해본다면 어떤 일[事], 즉 여기서 evidence를 관측하여 갱신하기 전 후의 $P(H)$ 주장에 관한 신뢰도 정도로 이해하면 될 것이다.

H'(어떤 사건이 발생했다는 가설)

H'(새로운 정보를 기반으로 어떤 사건이 발생했다는 가설)

사전 확률
(prior)

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

사후 확률
(posterior)

그림 1. 베이즈 정리는 사전확률과 사후확률간의 관계에 대해 설명하는 정리이다.

베이즈 정리의 의미와 의의

결론부터 말하자면 베이즈 정리는 새로운 정보를 토대로 어떤 사건이 발생했다는 주장에 대한 신뢰도를 갱신해 나가는 방법(a method to update belief on the basis of new information)이다.

베이즈 정리를 이해하기 어려웠던 이유

베이즈 정리를 이해함에 있어서 가장 먼저 정리해야 할 개념은 '확률'에 관한 관점이다.

베이즈 정리의 의미를 이해하기 어려운 이유 중 하나는 고등학교 수준까지의 확률론에서는 '전통적인 관점'으로 확률을 정의해오고 이해해왔기 때문이다.

여기서는 확률이라는 단어를 '주장에 대한 신뢰도'로 생각해보자.

이러한 관점은 확률에 대한 베이지안 주의(Bayesianism) 관점으로 볼 수 있다. 반면, 전통적인 확률관은 빈도주의(frequentism)이라고 볼 수 있다.

가령 동전의 앞면이 나올 확률이 50%라고 하면, 빈도주의자들은 100번 동전을 던졌을 때 50번은 앞면이 나온다고 해석하고, 베이지안 주의자들은 동전의 앞면이 나왔다는 주장의 신뢰도가 50%라고 보는 것.

처음 Bayes의 정리를 보면 아주 복잡하게 느껴지는데 곰곰이 생각해보면 어려운 내용이 아니라는 것을 알게 됩니다.

일단 먼저 기본적인 용어를 먼저 알아보면



① **사전확률** (prior probability)

⇒ 관측자가 이미 알고 있는 사건으로부터 나온 확률로 위의 베이지 정리에서는 $P(A_1), P(A_2), \dots, P(A_n)$ 을 의미함

② **우도** (likelihood probability)

⇒ 이미 알고 있는 사건이 발생했다는 조건하에 다른 사건이 발생할 확률로 위의 베이지 정리에서는 $P(B|A_1), P(B|A_2), \dots, P(B|A_n)$ 을 의미함

evidence

③ **사후확률** (posterior probability)

⇒ 사전확률과 우도를 통해서 알게되는 조건부 확률로 베이지 정리에서는 $P(A_k|B)$ 을 의미합니다

Bayes의 정리를 그림을 통해서 이해를 해보면

표본공간 S 는 서로소인 A_1, A_2, A_3 의 합집합으로 이루어 질 때

