

베이즈 정리

1. 건강 검진에서 양성일 경우 99%로 간암인 검사에서 양성 나왔다. 간암에 걸렸을 확률은 몇일까?

(전체 인구 중 간암에 걸린 환자는 0.1%)



$P(\text{양성}|\text{간암})$ 일 확률이 99% 인것 하지만 알아야 되는것은 $P(\text{간암}|\text{양성})$

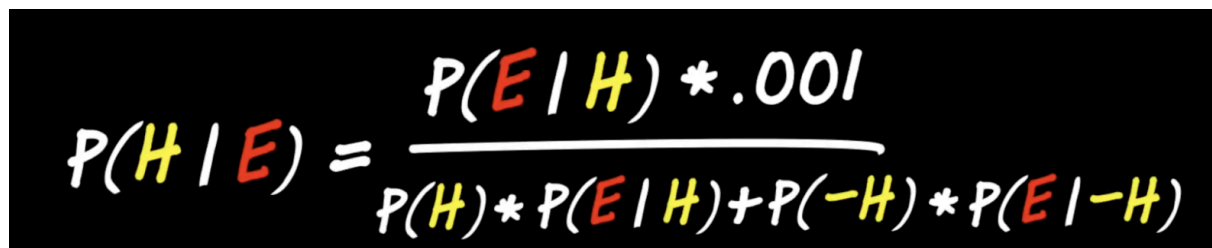
2. 출장을 위해 비행기를 타려고 합니다. 당신은 우산을 가져가야 하는지 알고 싶어 출장지에 사는 친구 3명에게 무작위로 전화를 하고 비가 오는 경우를 독립적으로 질문했습니다. 각 친구는 2/3로 진실을 말하고 1/3으로 거짓을 말합니다. 3명의 친구가 모두 “그렇습니다. 비가 내리고 있습니다”라고 말했습니다. 실제로 비가 내릴 확률은 얼마입니까?

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

H : 간암에 걸린 사건

E : 양성 나오는 사건

P(E) : 사전확률


$$P(H | E) = \frac{P(E | H) * .001}{P(H) * P(E | H) + P(-H) * P(E | -H)}$$

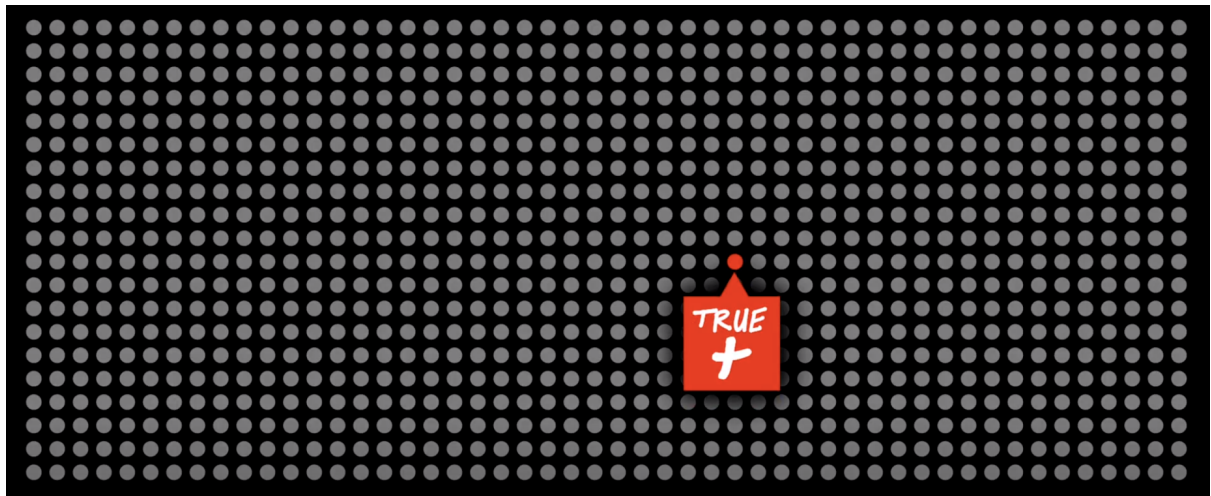
E : 양성 나오는 전체 사건은 간암에 걸린 경우와 아닌 경우의 합으로 나타낼수 있음.

$$P(H|E) = \frac{.99 * .001}{.001 * .99 + .999 * .01}$$

계산하면 약 9% 정도가 나옴.

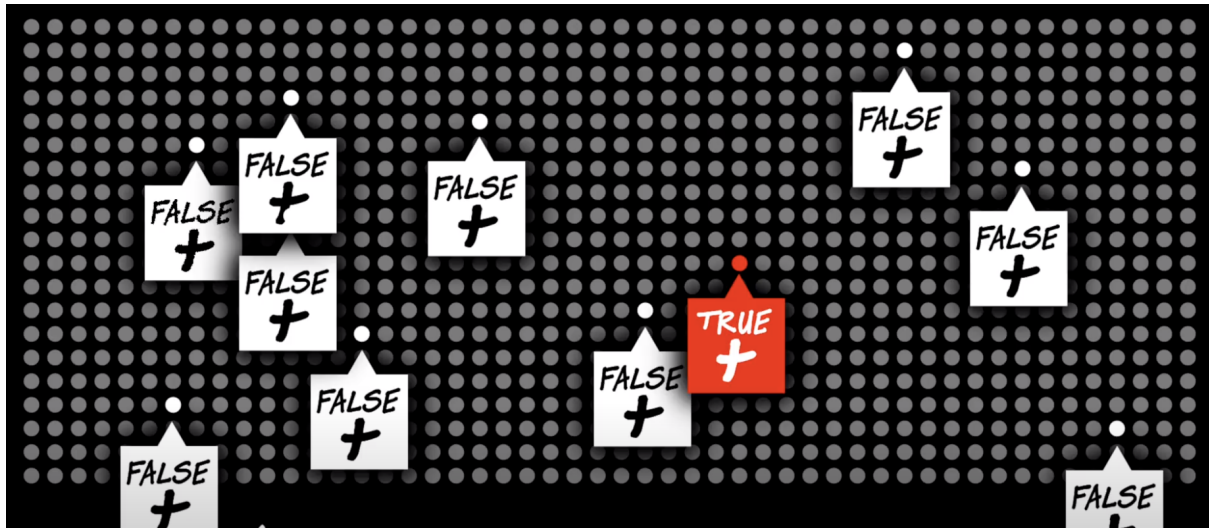
eg.

1000명의 사람이 있다고 하면,



사전확률에 따르면 그 중 한 사람만 간암에 걸렸을 것이다. (0.01%)

이 중 999명 모두 간암 검사를 받게 되면,



약 10명정도가 양성 반응이 나오게 된다 (99%)

그러면 양성반응이 나온 11명 중에서



1명만 간암이니 $1/11 = 0.09$ 9%가 나오게 된다.

베이즈 정리의 핵심은 사전확률을 가지고 사후확률을 계속 업데이트 할수 있다는 것이다.

예를들어 앞의 사람이 한번 더 검사를 했는데 또 양성인 나온다면, 이전의 사전확률 0.01%가

9%로 바뀌게 되고,

$$P(H | E) = \frac{0.0891}{0.0891 + .91 * .01}$$

이 사람이 병에 걸렸을 확률은 91%가 된다.

2. 배우는 이유

이처럼 베이즈 정리를 이용할 경우, 데이터를 토대로 사후 확률을 지속적으로 업데이트 해 나갈 수 있다. 우리가 마주치는 머신러닝 문제들도 어떤 가설을 설정하고 데이터(=관찰 결과)를 토대로 가설을 지속적으로 업데이트 해나가는 과정이기 때문에 인공지능의 이론적인 토대가 된다.