

MAP,MLE

1. Posterior

말 그대로 사후확률이다. 반대로 prior(사전확률)이 있고, 비교하는 것은 likelihood가 있다.



커튼에 비친 형상을 관찰할 수 있다고 하자. 이 때, 두 가지 경우를 생각 할 수 있다.

$P(\text{형상}|\text{철수})$, $P(\text{철수}|\text{형상})$

실제로 우리가 알고 싶은 것은 $P(\text{철수}|\text{형상})$ 이다.

이를 사후확률,posterior라고 부른다. 근데 실제로 이 확률은 우리가 알 수 없다.

그래서 근사할때 $p(\text{형상}|\text{철수})$ 를 쓴다. 또한 이를 likelihood라고 부를 수 있다.

- 형상이 철수 일 수 있다는 확률은 직관적으로 이해가 될 수 있는데, 후자는 그러지 않을 수 있어요.
- $P(\text{형상}|\text{철수})$ 의 의미는 철수가 여러가지 형상을 가질 수 있는데, 동물 형상, 사람 형상, 비행기 형상 등등.

- 이러한 형상들 중, 커튼에 비쳐져 있는 형상일 확률을 뜻하는 겁니다.

따라서, 사후확률은 우리가 잘 아는 베이즈 정리로 근사할 수 있다.

$$P(\text{철수}|\text{형상}) = \frac{P(\text{형상}|\text{철수})P(\text{철수})}{P(\text{형상})}$$

분모에 해당하는 $p(\text{형상})$ 의 값은 분자에 무엇이 오던간에 변하지 않는 상수에 해당한다.
결국,

$$P(\text{철수}|\text{형상}) \propto P(\text{형상}|\text{철수})P(\text{철수})$$

이와 같은 형식으로 근사하게 되고 ,이를 수학적 표현으로 나타내면 다음과 같다.

$$P(\text{posterior}) \propto \text{likelihood} \times P(\text{prior})$$

위의 식에서도 구할 수 있듯이, $p(\text{철수}), p(\text{영희})$ 와 같은 우리가 이미 알고 있고 정의된 확률을 **Prior(사전확률)**이라고 한다.

따라서, 베이지안 확률은 posterior를 likelihood와 prior를 활용해 구해주는 방법이라고 할 수 있다.

👁 **likelihood**: $p(z|x)$, 어떤 모델에서 해당 데이터(관측값)이 나올 확률

👁 **사전확률(prior probability)**: $p(x)$, 관측자가 관측을 하기 전에 시스템 또는 모델에 대해 가지고 있는 선험적 확률. 예를 들어, 남녀의 구성비를 나타내는 $p(\text{남자})$, $p(\text{여자})$ 등이 사전 확률에 해당한다.

👁 **사후확률(posterior probability)**: $p(x|z)$, 사건이 발생한 후(관측이 진행된 후) 그 사건이 특정 모델에서 발생했을 확률

2. MLE와 MAP

- MLE(Maximum likelihood)
- MAP(Maximum a Posterior)

먼저, MLE는 사람에게서 그러한 형상이 나올 확률 $P(\text{형상}|\text{철수})$ 와 $P(\text{형상}|\text{영희})$ 중, 단순히 큰 likelihood값을 고르는 것이다.

예를 들어서, 철수와 영희에게 총 1000번의 형상을 취하라고 했을 때, 내가 원하는 형상이 각각 0.4, 0.6의 확률로 나왔다면 우리는 철수에게서 우리가 원하는 형상이 나올 확률을 0.4라고 정할 것이다. 이렇게 정하는 방법이 MLE.

MAP는 반대로 어떤 형상을 보았는데 그 형상을 취하고 있는 사람이 철수 $P(\text{철수}|\text{형상})$, 영희 $P(\text{영희}|\text{형상})$ 를 비교해서 둘 중 큰 값을 갖는 클래스를 선택하는 방법이다. 즉, Posterior를 최대화 시키는 방법이며, 이때 베이즈 정리가 사용되게 된다.

이 두가지의 큰 차이점은 prior를 곱하냐 안곱하냐이다.

MAP에서 데이터에 대한 즉, $P(\text{철수})$, $P(\text{영희})$ 와 같은 prior 확률이 같다고 가정하면 MLE와 같은 수식이 나오게 된다. 그렇다면 MLE와 다르게 우리에게 필요한 것은 어떤 사람이 철수일 확률을 알아야 한다.

기본적으로 A라는 사람이 보낸 메일에 대해서 스팸인지 기본 메일인지를 구하면 MLE가 되겠지만, 우리가 전체메일 중에서 스팸메일과 기본 메일의 비율을 알고 있다면, 더 정확한 방법인 MAP를 쓸 수 있다는 것이다.

그래서 두 방법 사이의 정확성을 따져보면 MAP가 정확하다. 하지만 현실에서는 MLE를 쓸 때도 많다. 왜냐하면 클래스만의 확률, 즉 사전확률을 구하는게 어렵기때문!

Reference

<https://hwiyong.tistory.com/27>