Maximum A Posteriori Estimation (MAP)

수업 목표

이번 수업의 핵심:

- Likelihood, Prior, Posterior의 개념
- Maximum A Posteriori Estimation의 이해
- MLE와 MAP의 비교

핵심 개념

- Likelihood, Prior, Posterior
- Maximum A Posteriori Estimation

윷 던지기 예제

베이지안 관점의 기계학습 모델링에서

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} P(X|\theta)\cancel{P(\theta)}$$

Prior를 상수로 가정

$\Rightarrow$ 어떤 θ도 가질 수 있다고 생각하는 관점

위 내용에서 $P(\theta)$는 상수로 가정했지만, 실제론 그렇지 않을 것

- 우리는 윷의 앞면이 나올 확률이 대략 0.5임을 사전에 알고 있음
- 즉, 직관적으로 $P(\theta = 0.5) > P(\theta = 0.001)$

그러나, 이 Prior를 정확히 알 방법은 없음

- 그래도 $P(\theta)$를 적당히 가정하여 $\hat{\theta}$ 결과를 보완할 수 있을까?

Maximum A Posteriori Estimation

베이지안 관점의 기계학습:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \underbrace{P(\theta|X)}_{\substack{\text{Posterior} \\ \text{사후 분포}}} = \underset{\theta}{\operatorname{argmax}} \underbrace{P(X|\theta)}_{\text{Likelihood}} \underbrace{P(\theta)}_{\substack{\text{Prior} \\ \text{사전 분포}}}$$

- 관측된 X에서 최적의 θ를 찾아야 하므로, Posterior를 최대화 하는 것이 원래 목적

용어 정리:

- Posterior distribution, 사후 분포 $P(\theta|X)$: 데이터를 관측한 후의 θ 분포
- Prior distribution, 사전 분포 $P(\theta)$: 데이터를 관측하기 전의 θ 분포

Maximum A Posteriori Estimation (MAP):

- Likelihood에 Prior를 곱하여 Posterior를 최대화
- 하지만 Prior를 알 방법이 없으므로 가정하여 곱해줄 수 밖에 없음

Maximum A Posteriori Estimation: 윷 던지기 예제

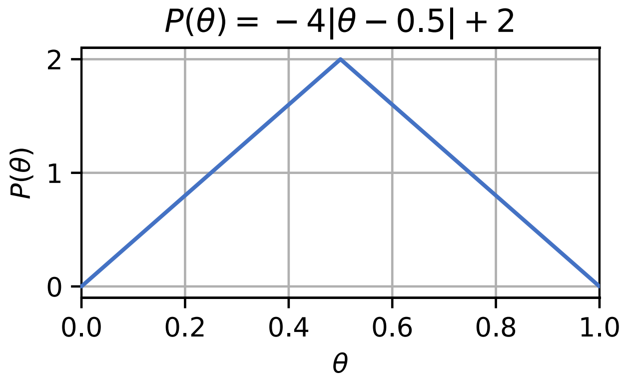
- 하나의 윷을 10회 던졌을 때 결과가 다음과 같이 나왔다
  - HHTHTTHHHT (H: Head 앞면, T: Tail 뒷면)

윷의 수학적 모델:

$$P(x|\theta) = \begin{cases} \theta, & x = \mathrm{H} \\ 1-\theta, & x = \mathrm{T} \end{cases}$$

$$P(X|\theta) = P(x_1, x_2, \dots, x_n|\theta) = P(x_1|\theta) \times P(x_2|\theta) \times \cdots \times P(x_n|\theta)$$

- 사전 분포 $P(\theta) = -4|\theta - 0.5| + 2$



앞면이 나올 확률이 0.5 주변에 있을 가능성이 크니까
0.5일 확률이 가장 높은 임의의 분포로 가정해보자
(다른 원하는 사전 분포를 가정해도 무방함)

Maximum A Posteriori Estimation: 윷 던지기 예제

MAP 관점에서 최적의 $\boldsymbol{\theta}$는?

$$\begin{aligned}\hat{\theta}_{MAP} &= \underset{\theta}{\operatorname{argmax}}\, P(X|\theta)P(\theta) \\ &= \underset{\theta}{\operatorname{argmax}}\, \theta^6(1-\theta)^4(-4|\theta-0.5|+2)\end{aligned}$$

- $\underset{\theta}{\operatorname{argmax}}\, \theta^6(1-\theta)^4(-4|\theta-0.5|+2)$를 찾기 위해 미분

$$\frac{d}{d\theta}\theta^6(1-\theta)^4(-4\theta) = 4\theta^6(11\theta-7)(1-\theta)^3 = 0 \qquad \theta \in [0,0.5]$$

$$\frac{d}{d\theta}\theta^6(1-\theta)^4(-4\theta+4) = -4\theta^5(11-6\theta)(1-\theta)^4 = 0 \qquad \theta \in [0.5,1]$$

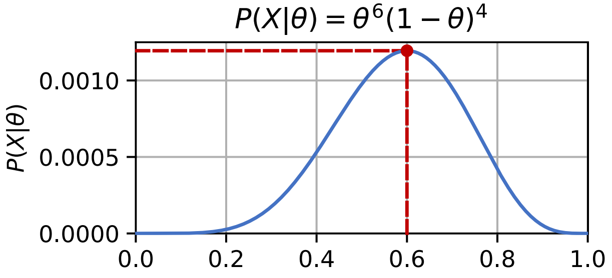
대입하면, $\theta = 0, 1$인 지점은 극소값, $\theta = \frac{7}{11}$은 범위 밖, $\theta = \frac{6}{11}$ 인 지점이 극대

$$\therefore \hat{\theta}_{MAP} = \frac{6}{11}$$

Maximum A Posteriori Estimation: 윷 던지기 예제

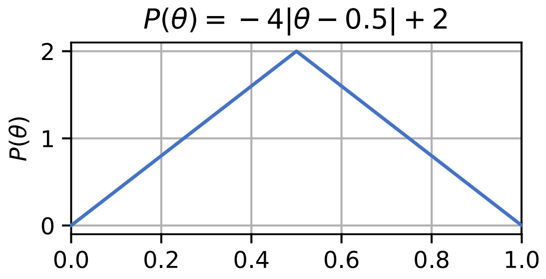
- Posterior

Likelihood



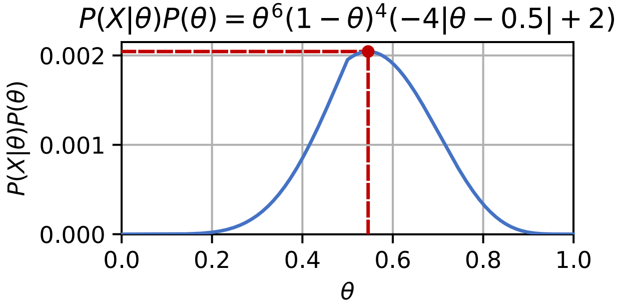
×

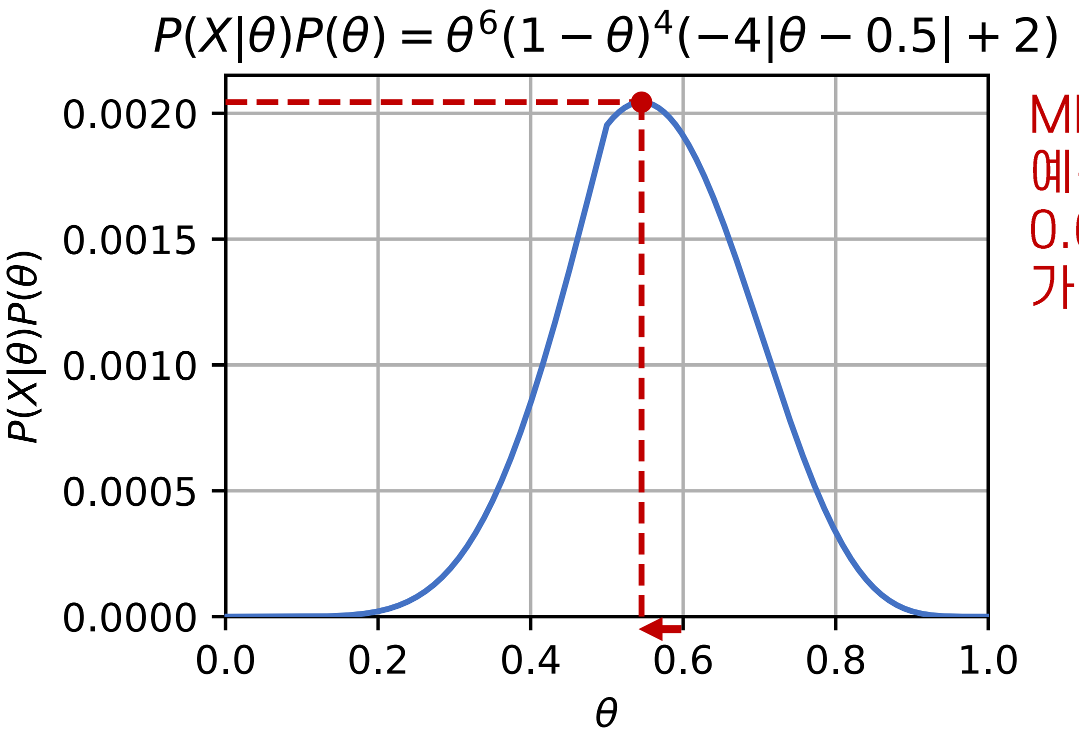
Prior



=

Posterior





MLE에서
예측했던
0.6보다 0.5에
가까워짐

MLE vs MAP

MAP의 장점: 적절한 사전분포의 모델링은 과적합으로부터 강건해질 수 있음

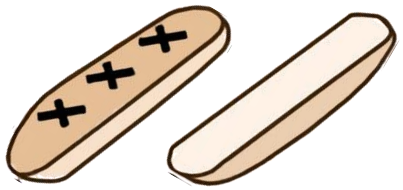
- 시행 횟수 / 학습 데이터 개수가 적은 경우, 적절한 Prior를 통해 보정이 가능
- 만약 윷을 3번만 던져서 모델링하는데, 모두 H가 나온 경우,

$$\hat{\theta}_{MLE} = 1, \qquad \hat{\theta}_{MAP} = 0.75$$ (앞의 예제와 같은 Prior 사용)

MAP의 단점: 과한 (잘못된) 사전 분포의 가정은 모델링에 방해가 될 수 있음

- 윷은 볼록하게 생겨서 앞면이 나올 확률은 실제론 0.5보다 작은 곳에 있을 것

상황에 맞게 적절한 사전 분포의 가정이 필요!



Prior를 고려하지 않는 MLE는 단순히 $\hat{\theta}_{MLE} = 0.6$으로, H가 나왔던 확률과 같게 모델링

$$\hat{\theta}_{MLE} = 0.6$$

윷의 특성에 맞추어 Prior를 0.5에 더 높이 준 MAP는 MLE의 예측 값과 0.5의 중간에서 모델링

$$\hat{\theta}_{MAP} = \frac{6}{11} = 0.5454\ldots$$

요약

- 윷 던지기 예제를 통한 Maximum A Posteriori Estimation의 이해
- Bayes' Rule을 이용한 Posterior의 Prior & Likelihood 꼴 변형
- 사전 분포를 정의함으로써 MLE와 대비되는 MAP의 장단점