Maximum Likelihood Estimation (MLE)

수업 목표

이번 수업의 핵심:

- Maximum Likelihood Estimation의 필요성
- 베이지안 관점에서의 기계학습
- MLE에 기반한 최적의 모델 탐색

핵심 개념

- Likelihood, Prior
- Maximum likelihood estimation
- Log-likeihood

윷이 앞면이 나올 확률은?

- 윷이 앞면이 나올 확률은 거의 0.5이지만 완전히 0.5는 아닐 것

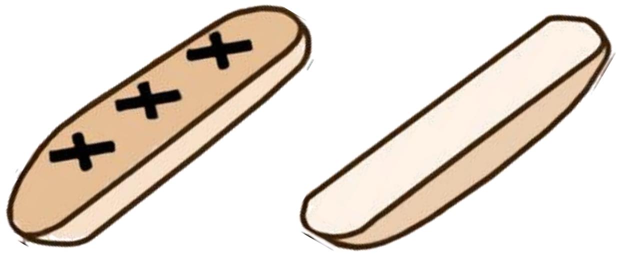
→ 정확한 확률을 알기 위하여 윷을 던져서 앞면이 나올 확률을 알아보자

- 하나의 윷을 10회 던졌을 때 결과가 다음과 같이 나왔다

HHTHTTHHHT (H: Head 앞면, T: Tail 뒷면)

- 위 실험에 기반해서, 윷이 앞면이 나올 확률은?

$$\frac{\text{앞면이 나온 횟수}}{\text{전체 시행 횟수}} = \frac{6}{10} = 0.6$$



- **이때, 윷이 앞면이 나올 확률을 0.6이라 할 수 있을까?**

윷 던지기 예제

- 윷의 확률적 모델 (앞면이 나올 확률을 θ로 가정):

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

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

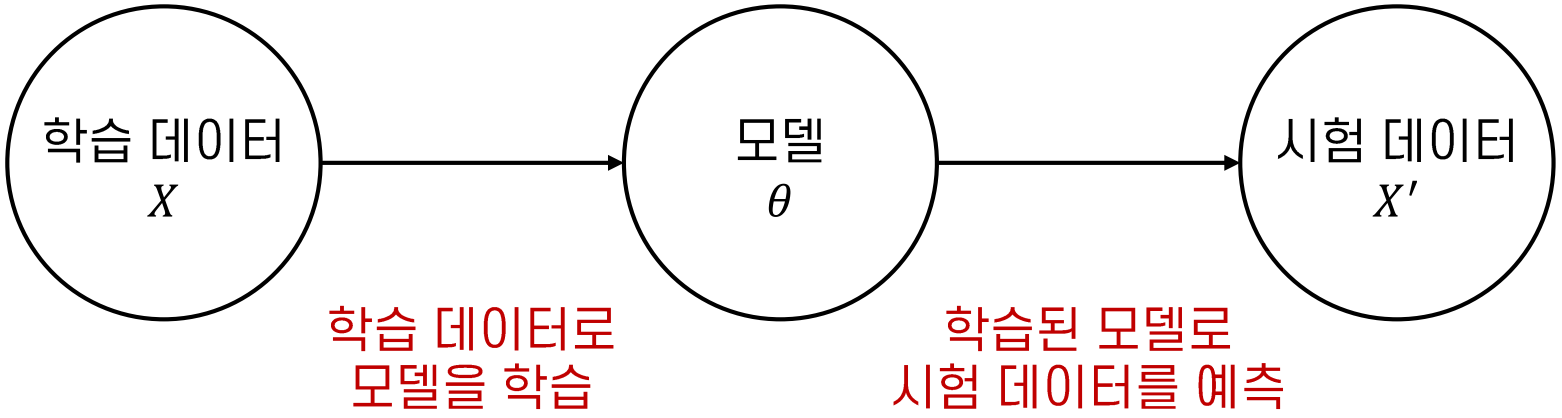
- θ를 알고 있다면:
  - 윷을 10번 던져 X = HHTHTTHHHT가 나올 확률을 계산할 수 있음
  - 윷을 던졌을때 나오는 결과에 해당하는 X의 분포를 그려볼 수 있음 → $P(X|\theta)$

그러나, 우리가 궁금한 것은 X = HHTHTTHHHT가 나왔을 때의 θ값: $P(\theta|X)$

- θ는 일종의 분포를 나타냄
- 이때, X가 주어졌을 때 가장 확률 값이 높은 θ를 찾는 방법은?

베이지안 관점의 기계학습

- 학습 데이터 X로 모델 θ를 학습하여 시험 데이터 X'에서 잘 맞추는 것이 목표
- 베이지안 관점의 기계 학습: X, X' 뿐만 아니라 θ 또한 확률 변수로 생각



베이지안 관점의 기계학습

목표: 학습 데이터 X가 주어질 때 이를 가장 잘 나타낼 수 있는 $\hat{\theta}$를 찾는 것

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

- X = HHTHTTHHHT가 나왔을 때 윷을 모델링할 가장 타당한 θ를 찾자!

- 그러나 $P(\theta|X)$를 모델링하는 함수를 찾기는 어려움

→ 대신, Bayes' Rule을 이용하여 $P(X|\theta)$를 활용하자!

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

$$= \underset{\theta}{\operatorname{argmax}}\, \underbrace{P(X|\theta)}_{\text{Likelihood}}\underbrace{P(\theta)}_{\text{Prior}}$$

$P(X)$는 θ와 관련이 없음

Maximum Likelihood Estimation

Likelihood의 정의:

$$\text{Likelihood} \coloneqq P(X|\theta)$$

- 해석1: θ를 가정했을 때 학습 데이터 X가 얼마나 잘 모델링 되었는가?
- 해석2: 베이지안 관점의 기계학습 모델링에서

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

$P(\theta)$를 상수로 가정 /
Prior를 Uniform distribution으로 가정

$\Rightarrow$ 어떤 θ도 가질 수 있다고 생각하는 관점

  - θ가 0.1일수도 0.7일수도 있으며, 딱히 우리의 편견이 들어가지 않음
- Likelihood가 높다 $\Rightarrow$ 학습 데이터를 잘 모델링 했다
- **Maximum Likelihood Estimation (MLE)**: Likelihood가 최대인 θ를 찾기

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}}\, P(X|\theta)$$

Maximum Likelihood Estimation: 윷 던지기 예제

MLE 관점에서 최적의 $\boldsymbol{\theta}$는?

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

$$\begin{aligned} \hat{\theta}_{MLE} &= \operatorname*{argmax}_{\theta} P(X|\theta) = \operatorname*{argmax}_{\theta} \prod_{i=1}^{10} P(x_i|\theta) \\ &= \operatorname*{argmax}_{\theta} \theta\theta(1-\theta)\theta(1-\theta)(1-\theta)\theta\theta\theta(1-\theta) \\ &= \operatorname*{argmax}_{\theta} \theta^6(1-\theta)^4 \end{aligned}$$

- $\operatorname*{argmax}_{\theta} \theta^6(1-\theta)^4$를 찾기 위해 $\theta^6(1-\theta)^4$를 미분
  - 미분하여 0이 되는 지점이 극대/극소값

$$\begin{aligned} \frac{d}{d\theta}\theta^6(1-\theta)^4 &= 6\theta^5(1-\theta)^4 - 4\theta^6(1-\theta)^3 \\ &= \theta^5(1-\theta)^3(6(1-\theta) - 4\theta) \\ &= \theta^5(1-\theta)^3(6-10\theta) \end{aligned}$$

- 대입해보면, $\theta = 0, 1$인 지점은 극소값, $\theta = 0.6$인 지점이 극대값

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

Log-likelihood

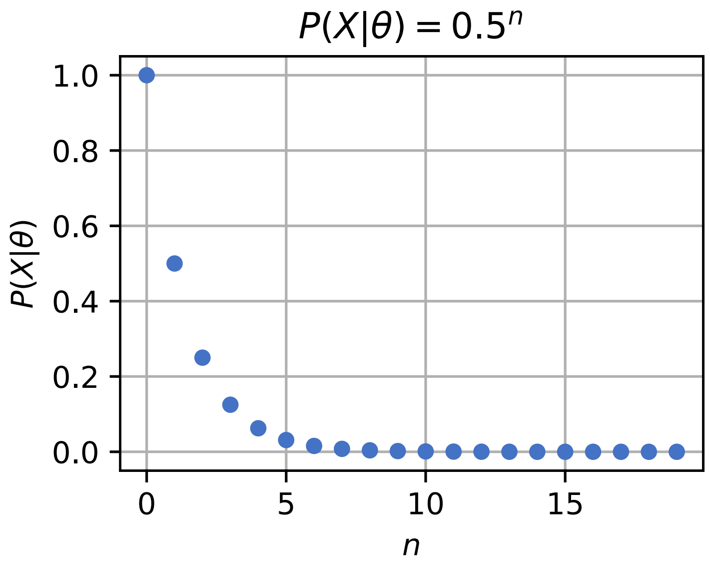
- 데이터에 대한 Likelihood를 나타내려면 데이터 수만큼 확률값을 곱해야 함

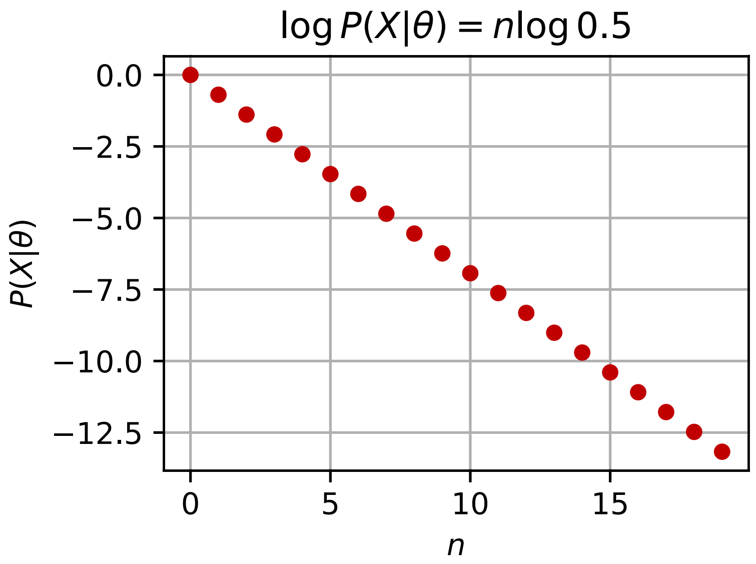
$$P(X|\theta) = P(x_1|\theta)P(x_2|\theta) \cdots P(x_N|\theta)$$

- $P(x_i|\theta) \leq 1$을 반복적으로 곱하면 $P(X|\theta)$가 기하급수적으로 0에 근접
- Log-likelihood의 정의:

왜 0에 근접하면 안되나요?
너무 작은 수에 대한 연산은 컴퓨터에서 정확하지 않음

$$\log P(X|\theta) = \log P(x_1|\theta)P(x_2|\theta) \cdots P(x_N|\theta) = \log P(x_1|\theta) + \cdots + \log P(x_N|\theta)$$





요약

- 윷 던지기 예제를 통한 Maximum likelihood estimation의 이해
- 베이지안 관점에서의 Likelihood 정의와 이에 대한 해석
- MLE 관점에서 주어진 데이터의 확률이 최대가 되는 모델 탐색 예시