선수 지식 - 통계
최대 가능도 추정

최대 가능도 추정 | 딥러닝의 기초가 되는 확률 개념 알아보기

강사 나동빈

선수 지식 - 통계

최대 가능도 추정

최대 가능도 추정(Maximum Likelihood Estimation)

- 인터넷에서 데이터 x를 N개 수집했다.
- 이러한 N개의 데이터가 정규 분포를 따른다고 가정하고, 정규 분포를 추정(estimation)해보자.
- **모수 추정 문제**: 표본 값 x에 대해서는 알고 있지만, 모수 θ를 모르는 상황
- 모멘트 방법 말고, **최대 가능도 추정**을 사용할 수 있다.

최대 가능도 추정(Maximum Likelihood Estimation)

- 이론적으로(수식적으로) **가장 가능성이 높은** 모수(parameter)를 찾는 방법이다.
- 기본적으로 모든 추정 방법 중에서 가장 널리 사용되는 방법 중 하나다.
- 여러 가지 확률 분포 X에 대한 확률 함수를 다음과 같이 표현하자.

$$p(x;\theta)$$

- 이때 x는 확률 분포가 가질 수 있는 실수 형태의 값이다.
- θ와 x 모두 스칼라이거나 벡터이다.

최대 가능도 추정(Maximum Likelihood Estimation)

- 가지고 있는(관측된) 데이터 x를 토대로 모수 θ를 찾는 문제로 이해할 수 있다.
- 즉, 확률 밀도 함수에서 **모수를 변수로 간주**하는 것이다.
- **가능도 함수**: $L(\theta; x) = p(x; \theta)$
- 추정하고자 하는 확률 분포에 따라 가능도 함수를 다르게 정의할 수 있다.

1) **베르누이 확률 분포**를 추정하는 경우

$\rightarrow \theta = \mu$

2) **정규 분포**를 추정하는 경우

$\rightarrow \theta = (\mu, \sigma^2)$

최대 가능도 추정(Maximum Likelihood Estimation)

- 최대 가능도 추정은 다음과 같은 문제를 해결하는 것이 목표다.

$$\hat{\theta}_{MLE} = \arg\max_{\theta} L(\theta; x)$$

→ 가지고 있는 정보를 토대로, **가능도(likelihood)**를 최대로 만드는 $parameter$를 찾는다.

정규분포 추정 예시

- 정규분포의 확률밀도함수는 다음과 같다. (μ와 σ가 상수로 주어진 상황)

→ 모수 θ를 알고 있으며, 적분했을 때 면적은 항상 1이다.

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

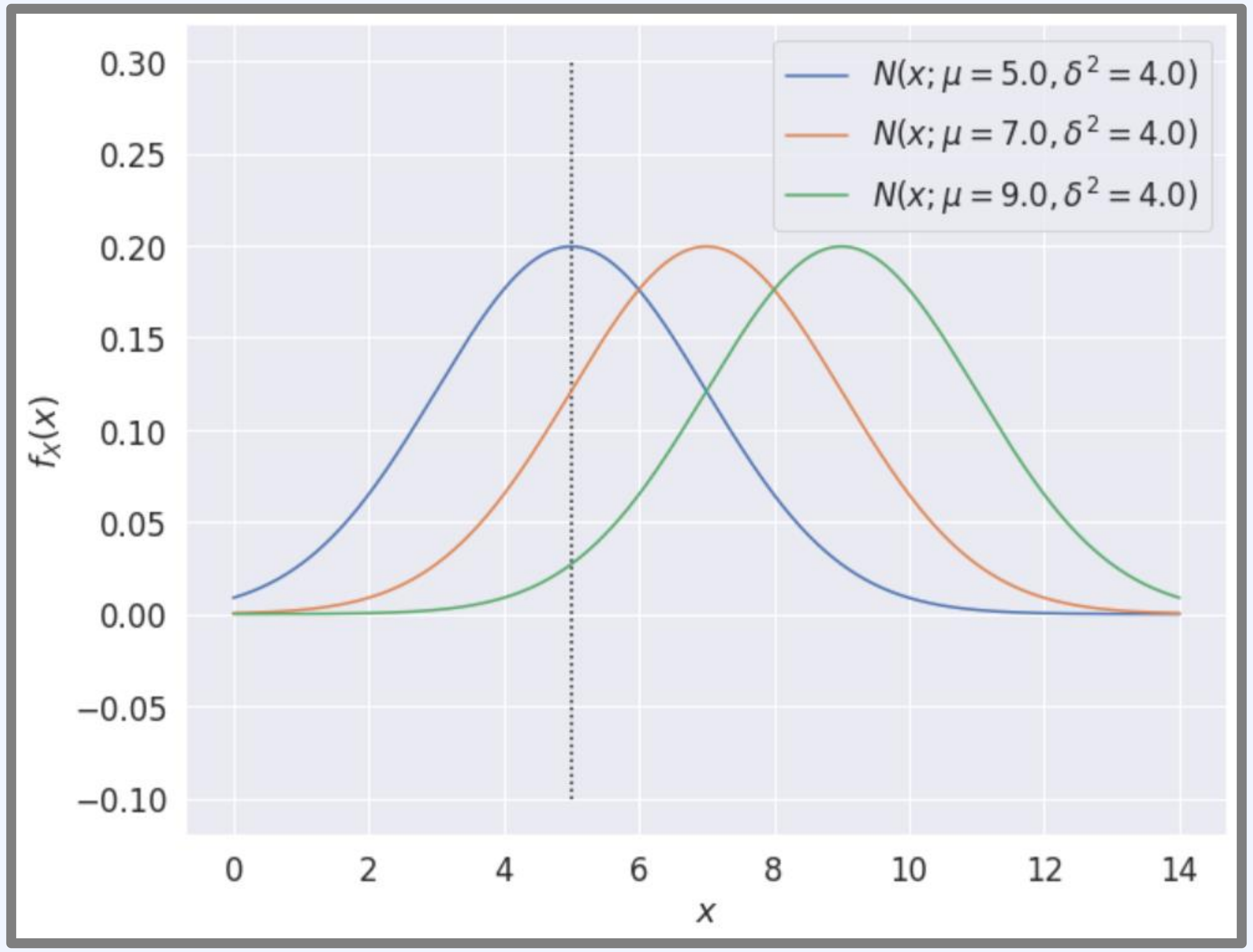
- **가능도 함수**는 다음과 같다. (x가 상수로 주어진 상황)

→ 데이터 x를 알고 있으며, 적분했을 때 면적은 1이 아닐 수 있다.

$$L(\mu, \sigma^2; x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

정규분포 추정 예시

- **분산(variance)**이 4로 알려져 있고, 값이 5인 하나의 데이터를 가지고 있다고 해보자.
- 이때 [5, 7, 9] 중에서 어떤 값이 가장 **평균(mean)**에 적합할까?



여러 개의 데이터가 있는 경우 가능도 추정

- N개의 데이터 $\{x_1, x_2, \dots, x_N\}$를 가지고 있는 상황을 고려해 보자.
- 각 표본 데이터는 같은 확률 분포에서 나온 **독립**적인 값이다.
- **독립적**: 1번째 샘플링에서 x_1이 나온 것이 2번째 샘플링의 값 x_2를 결정하지 않는다.
- 따라서 N개의 데이터가 동시에 나올 결합 확률 밀도 함수는 다음과 같다.

$$L(\theta; x_1, \dots x_N) = p(x_1, \dots, x_N; \theta) = \prod_{i=1}^{N} p(x_i; \theta)$$

여러 개의 데이터가 있는 경우 가능도 추정

- 정규 분포에서 다음의 4개의 데이터를 얻은 상황을 가정하자.

$$\{-5, 0, 3, 10\}$$

$$L(\theta; x_1, x_2, x_3, x_4) = \mathcal{N}(x_1, x_2, x_3, x_4; \theta) = \mathcal{N}(x_1; \theta) \cdot \mathcal{N}(x_2; \theta) \cdot \mathcal{N}(x_3; \theta) \cdot \mathcal{N}(x_4; \theta)$$

$$= \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(-5-\mu)^2}{2\sigma^2}\right) \cdot \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(0-\mu)^2}{2\sigma^2}\right) \cdot \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(3-\mu)^2}{2\sigma^2}\right) \cdot \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(10-\mu)^2}{2\sigma^2}\right)$$

$$= \frac{1}{\left(\sqrt{2\pi\sigma}\right)^4} \exp\left(-\frac{(-5-\mu)^2 + (0-\mu)^2 + (3-\mu)^2 + (10-\mu)^2}{2\sigma^2}\right)$$

$$= \frac{1}{\left(\sqrt{2\pi\sigma}\right)^4} \exp\left(-\frac{4\mu^2 - 16\mu + 134}{2\sigma}\right) = \frac{1}{\left(\sqrt{2\pi\sigma}\right)^4} \exp\left(-\frac{4(\mu-2)^2 + 118}{2\sigma^2}\right)$$

여러 개의 데이터가 있는 경우 가능도 추정

- 결과적으로 가능도(likelihood) 함수는 다음과 같다.

$$L(\theta; x_1, \dots x_N) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^4} \exp\left(-\frac{4(\mu-2)^2+118}{2\sigma^2}\right)$$

- σ 값에 상관없이 $\mu = 2$일 때 항상 최댓값을 가진다.

정규 분포의 최대 가능도 추정

- 정규 분포의 최대 가능도 함수를 다음과 같이 정리할 수 있다.

$$L(\mu; x_1, \dots x_N) = p(x_1, \dots, x_N; \mu) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right)$$

$$\because \log L = \log p(x_1, \dots, x_N; \mu) = \sum_{i=1}^{N} \left\{-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x_i - \mu)^2}{2\sigma^2}\right\}$$

$$= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - \mu)^2$$

정규 분포의 최대 가능도 추정

- 결과적으로 $\log(\cdot)$ 함수를 취해 정리된 가능도 함수는 다음과 같다.

$$-\frac{N}{2}\log(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2$$

- 이를 μ와 σ^2로 미분한 값이 0일 때, 가능도(likelihood) 값이 최대가 된다.

$$\frac{\partial \log L}{\partial \mu}=\frac{\partial}{\partial \mu}\left\{\frac{N}{2}\log(2\pi\sigma^2)+\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2\right\}=0$$

$$\frac{\partial \log L}{\partial \sigma^2}=\frac{\partial}{\partial \sigma^2}\left\{\frac{N}{2}\log(2\pi\sigma^2)+\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2\right\}=0$$

정규 분포의 최대 가능도 추정

- 식을 전개하면, 해는 다음과 같다.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2$$

→ 정규 분포의 **평균(μ)**은 표본 평균과 같고, **분산(σ^2)**은 표본 분산과 같다.