Maximum Likelihood Estimation (MLE)

Jerome Chou

February 24, 2025

Introduction to MLE

Maximum Likelihood Estimation (MLE) is a method for estimating the parameters of a statistical model by maximizing the likelihood function.
Given a dataset $X = \{x_1, x_2, \ldots, x_n\}$, and a parameterized probability distribution $p(x|\theta)$, MLE finds θ such that:

$$\hat{\theta} = \arg\max_{\theta} L(\theta; X) \tag{1}$$

where the likelihood function is:

$$L(\theta; X) = \prod_{i=1}^{n} p(x_i|\theta) \tag{2}$$

Log-Likelihood Function

Since the likelihood function involves a product, we often take the natural logarithm to simplify calculations:

$$\ell(\theta; X) = \log L(\theta; X) = \sum_{i=1}^{n} \log p(x_i|\theta) \tag{3}$$

MLE then simplifies to:

$$\hat{\theta} = \arg\max_{\theta} \ell(\theta; X) \tag{4}$$

Example: MLE for Gaussian Distribution

Assume $X \sim \mathcal{N}(\mu, \sigma^2)$. The probability density function is:

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{5}$$

The log-likelihood function is:

$$\ell(\mu, \sigma^2; X) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{6}$$

Taking derivatives and solving:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 \tag{7}$$

MLE in Machine Learning

MLE is widely used in machine learning applications, including:

- Logistic Regression: Maximizing the likelihood for binary classification.
- Gaussian Mixture Models (GMM): Estimating parameters using Expectation-Maximization (EM).
- Neural Networks: Training models by maximizing the likelihood of observed labels.

The loss function in many ML models is derived from the negative log-likelihood.

Conclusion

- MLE is a fundamental technique for parameter estimation.
- Log-likelihood simplifies optimization.
- It is widely used in statistics and machine learning.

Key Takeaway: MLE provides a principled way to estimate model parameters by maximizing the likelihood of observed data.