Maximum Likelihood Estimation

Your Name

May 22, 2025

Outline

Why Maximum Likelihood Estimation?

- MLE is a fundamental method in statistics and machine learning.
- It provides a way to estimate parameters of a statistical model.
- Widely used in various applications: regression, classification, etc.
- Helps in understanding the underlying data distribution.

The Problem

- Given a dataset $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$.
- We want to estimate the parameters θ of a probability distribution $P(X|\theta)$.
- The goal is to find θ that maximizes the likelihood function:

$$L(\theta|\mathbf{X}) = P(\mathbf{X}|\theta)$$

Intuitive Understanding

- Likelihood measures how well the model explains the observed data.
- Higher likelihood indicates a better fit of the model to the data.
- MLE finds the parameter values that make the observed data most probable.

Likelihood Function

- For independent and identically distributed (i.i.d.) samples:

$$L(\theta|\mathbf{X}) = \prod_{i=1}^{n} P(x_i|\theta)$$

- Log-likelihood:

$$\ell(\theta|\mathbf{X}) = \log L(\theta|\mathbf{X}) = \sum_{i=1}^{n} \log P(x_i|\theta)$$

MLE for Bernoulli Distribution

- Let $X_i \sim \text{Bernoulli}(\theta)$, where θ is the probability of success.
- Likelihood function:

$$L(\theta|\mathbf{X}) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

- Log-likelihood:

$$\ell(\theta|\mathbf{X}) = n\log(\theta) + (n-k)\log(1-\theta)$$

where $k = \sum_{i=1}^{n} x_i$.

Maximizing the Log-Likelihood

- Differentiate the log-likelihood:

$$\frac{d\ell}{d\theta} = \frac{k}{\theta} - \frac{n-k}{1-\theta}$$

- Set the derivative to zero:

$$\frac{k}{\theta} = \frac{n-k}{1-\theta}$$

- Solve for θ:

$$\hat{\theta} = \frac{k}{n}$$

Properties of MLE

- Consistency: $\hat{\theta} \xrightarrow{P} \theta$ as $n \to \infty$.
- Asymptotic Normality: $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I(\theta)^{-1})$.
- Efficiency: MLE achieves the Cramér-Rao lower bound.

MLE for Normal Distribution

- Let $X_i \sim \mathcal{N}(\mu, \sigma^2)$.
- Likelihood function:

$$L(\mu, \sigma^2 | \mathbf{X}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

- Log-likelihood:

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

Maximizing Log-Likelihood for Normal Distribution

- Differentiate with respect to μ and σ^2.
- Set derivatives to zero:

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

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

Applications of MLE

- Parameter estimation in regression models.
- Estimating distributions in anomaly detection.
- Used in various machine learning algorithms (e.g., Gaussian Mixture Models).

Challenges

- MLE can be sensitive to outliers.
- May not exist for some distributions.
- Computationally intensive for complex models.

Conclusion

- MLE is a powerful method for parameter estimation.
- Understanding its properties is crucial for effective application.
- Future work: Explore Bayesian estimation as an alternative.