Maximum Likelihood Estimation

Maathangi S

SOC'24 - Statistical Methods in Image Segmentation

June 2024

1 Introduction

MLE (Maximum Likelihood Estimation) is one of the primary ways/algorithms for estimating the parameters for a large amount of data given to us.

This algorithm aims to find parameters θ that fit the data most likely. The data from which the parameters will be estimated is n iid samples: $X_1, X_2, ..., X_n$.

2 Likelihood

Since the data is iid, they'll have the same distribution. We denote this PMF (for discrete data) or PDF (for continuous data) as $f(X|\theta)$. This particular notation is useful and best suited as the value of the function changes as we change the parameters, hence making it dependent on θ.

The likelihood of the data is the product of the likelihood of each point.

$$L(\theta) = \prod_{i=1}^{n} f(X_i|\theta)$$

3 Maximization of Likelihood

Our goal is to find the argument θ that maximizes the likelihood function.

$$\hat{\theta} = argmax_{\theta} L(\theta)$$

where argmax refers to the domain of the function that maximizes the function. An interesting property of argmax is that the argmax of a function is the same as the argmax of the log of the function. This is useful as the differentiation and other methods in log are comparatively simpler. We'll denote the log of the

Likelihood function as $LL(\theta)$.

$$LL(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(X_i|\theta)$$

Although the parameter can be calculated in a lot of ways, the commonly used method is to compute the first derivative.

4 Few examples

4.1 Bernoulli MLE

The parameter in the Bernoulli distribution is the probability p. Since the sample is iid distributed, $X_i \sim \text{Ber(p)}$ for each $i \in [1, n]$. Our goal is the find the value of p that maximizes the likelihood function.

$L(\theta) = \prod_{i=1}^{n} f(X_i|\theta) = \prod_{i=1}^{n} (1-p)^{1-X_i} p^{X_i}$

$LL(\theta) = \log L(\theta) = \sum_{i=1}^{n} (1 - X_i) \log(1-p) + X_i \log(p)$

$= (n - Y) \log(1-p) + Y \log(p)$ where $Y = \sum_{i=1}^{n} X_i$

By computing the first derivative by p of $LL(\theta)$ and equating it to 0, we get,

$\frac{n-Y}{1-p} = \frac{Y}{p} \implies p = \frac{Y}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$

4.2 Normal MLE

The parameters in normal distribution is $\theta = (\mu, \sigma^2)$. Since the sample is iid, $X_i \sim N(\mu, \sigma^2)$ for each $i \in [1, n]$. By using the above algorithm,

$LL(\theta) = \sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(X_i-\mu)^2}{2\sigma^2}} = \sum_{i=1}^{n} - \log(\sqrt{2\pi}\sigma) + \frac{(X_i-\mu)^2}{2\sigma^2}$

To estimate the parameters that most likely fit the data, we need to partially differentiate $LL(\theta)$ by μ, σ^2, and equate each equation to 0. By solving those, we get

$$\hat{\theta} = (\hat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n}, \frac{\sum_{i=1}^{n} (X_i - \hat{\mu})^2}{n})$$