Maximum Likelihood Estimation (MLE) for Linear Regression

Model Setup

We model the continuous target $y_i \in \mathbb{R}$ given features $\mathbf{x}_i \in \mathbb{R}^d$ using a linear function:

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$$

where $\boldsymbol{\beta} \in \mathbb{R}^d$ is the parameter vector and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ is Gaussian noise.

Likelihood Function

Given the Gaussian noise assumption, the conditional distribution of y_i is:

$$P(y_i \mid \mathbf{x}_i; \boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{2\sigma^2}\right)$$

The likelihood over n i.i.d. samples is:

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{2\sigma^2}\right)$$

Log-Likelihood

Taking the log of the likelihood:

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2$$

MLE Solution

To find the MLE, we maximize the log-likelihood. For fixed σ^2, this is equivalent to minimizing the sum of squared errors:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2$$

The solution is obtained by solving the normal equations. Let $X \in \mathbb{R}^{n \times d}$ be the design matrix and $\mathbf{y} \in \mathbb{R}^n$ the target vector. Then:

$$\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbf{y}$$

Variance Estimation

Once $\hat{\boldsymbol{\beta}}$ is found, the MLE of σ^2 is:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}})^2 = \frac{1}{n}\|\mathbf{y} - X\hat{\boldsymbol{\beta}}\|^2$$

Deriving Linear Regression from the Exponential Family

The Gaussian distribution belongs to the exponential family:

$$P(y \mid \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right)$$

For the normal distribution with mean μ and variance σ^2, we can write:

$$P(y \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

This fits the exponential family form with:

$$\theta = \mu, \quad b(\theta) = \frac{\theta^2}{2}, \quad \phi = \sigma^2, \quad c(y, \phi) = -\frac{y^2}{2\phi} - \frac{1}{2}\log(2\pi\phi)$$

In the GLM framework:

- The mean $\mu = \mathbb{E}[y] = \mathbf{x}^\top \boldsymbol{\beta}$
- The canonical link function is the identity: $g(\mu) = \mu$
- The linear predictor is $\theta = \mathbf{x}^\top \boldsymbol{\beta}$

Summary

- Linear regression assumes normally distributed errors.
- The likelihood is based on the Gaussian distribution.
- MLE reduces to minimizing the sum of squared errors.
- A closed-form solution exists via the normal equations.
- Linear regression is a GLM with identity link and normal noise.