Linear regression fitting via likelihood maximization

Statistics and Data Analysis

Nov 19, 2025

Table of Contents

Probabilistic model

A probabilistic model is a rule

$$\text{data} \sim \text{distribution}_{\text{parameters}}$$

trying to explain the data we see.

Distribution

A distribution is an absolute measure of the chances of getting individual data ranges through sampling.

Distribution examples

Examples

- uniform distribution – dx on $x \in [0, 1]$,
- normal (Gaussian) distribution – $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$ on $x \in \mathbb{R}$,
- exponential distribution – $\lambda e^{-\lambda x}$ on $x \geq 0$.

Sample probability?

The probability of having an exact data point is usually zero. How to compare parametrized distributions then?

Definition

Likelihood is a relative measure of how likely it is for a distribution to produce a given data point.

$$\text{likelihood} = \frac{\text{distribution}}{\text{baseline distribution}}$$

Baseline distribution is usually taken to be *improper* (e.g. dx for all $x \in \mathbb{R}$).

Normal distribution likelihood

Compared to the baseline distribution dx, the likelihood of a normal distribution $N(\mu, \sigma)$ producing a data point x is

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

Dataset likelihood

The likelihood of a distribution producing a dataset is defined as the product

$$\mathcal{L}(\mu, \sigma \mid \{x_i\}_{i=1}^{N}) = \prod_{i=1}^{N} \mathcal{L}(\mu, \sigma \mid x_i)$$

What are the distribution parameters of a normal distribution with an *unknown mean* μ and a *known standard deviation* σ explaining a dataset $\{x_i\}_{i=1}^{N}$ best?

Probabilistic description

The probabilistic model for a linear regression $Y \sim \beta_0 + \beta_1 X$ is

$$X, Y \sim e^{-(Y-\beta_0-\beta_1 X)^2} \, dX \, dY.$$