


[← Go Back to Machine Learning](#)
[☰ Course Content](#)

Maximum Likelihood and Empirical Risk Estimation

To establish the relationship between variables, it is required to calculate the parameters of the model. Now that we have understood the basics of supervised learning algorithms and linear regression, let us look at how to estimate the parameters of the linear regression model. We will see the intuition of two methods that are used to estimate these parameters:

Maximum Likelihood

Empirical Risk Minimization

Maximum Likelihood

Let (X_i, Y_i) be a sample in a certain dataset with n records. Consider $Y_i|X_i$ is an **event** that shows the occurrence of Y_i provided X_i has already occurred, i.e., given the input features X_i , the occurrence of Y_i .

In the entire sample, there are n such events. The **likelihood** of the event $Y_i|X_i$ is the probability of occurrence of this event.

In the context of linear regression, let us consider the relationship between Y_i and X_i which is as follows:

$$Y_i = \theta_0^* + \theta_1^* X_i + W_i$$

Where θ_0^* is the constant term in the equation, θ_1^* is the coefficient to the variable X_i , and W_i is the error term.

It is assumed that the events in the **sample are independent of each other** and the error terms **follow a normal distribution**, with a mean equal to the prediction and standard deviation σ . Keeping in mind the normal distribution, mathematically, the likelihood function is defined as follows:

$$P(Y_i|X_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left\{-\frac{(Y_i - \theta_0^* - \theta_1^* X_i)^2}{2\sigma^2}\right\}}$$

The above equation shows the likelihood of a single event. To estimate the collective likelihood of the entire sample of events, we need to combine the likelihood of all the individual events in the sample.

Now, as we assumed that the events in the sample are independent of each other, we can apply the **multiplication rule of probability**.

Hence, the likelihood of the entire sample of events can be given as follows:

$$L(\theta_0^*, \theta_1^*) = \prod_{i=1}^n P(Y_i|X_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left\{-\frac{(Y_i - \theta_0^* - \theta_1^* X_i)^2}{2\sigma^2}\right\}}$$

The above equation shows the combined probability of all the events to occur in the sample. To estimate the parameters, the above expression has to be maximized and the **maximum value** of the above expression is called the **maximum likelihood**.

Empirical Risk Minimization

If the error is high, then it can be considered that the algorithm is not making predictions close to the actual value, i.e., the prediction is at higher risk. Hence, the **higher the error**, the **higher the risk** and vice versa.

Empirical risk can be defined as the **collective error** made by the model over all the training records. During the estimation of the parameters of the model, we use the **principle of empirical risk minimization** where the empirical risk is minimized to get the model parameters. The expression of empirical risk can be given as follows:

$$Empirical\ Risk = \frac{1}{n} \sum_{i=1}^n W_i = \frac{1}{n} \sum_{i=1}^n (Y_i - \theta_0^* - \theta_i^* X_i)^2$$

A higher **empirical risk** is **equivalent to a lower likelihood** and vice versa.

◀ Previous

Next ▶

Proprietary content.©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

© 2023 All rights reserved.

[Help](#)