

[← Go Back to Data Analysis & Visualization](#)[☰ Course Content](#)

Kullback-Leibler (KL) Divergence

Sometimes, in machine learning, it is desirable to analyze the actual and observed probability distribution to quantify the difference in the distributions of a random variable. We calculate KL divergence to measure that difference.

First, let us understand what divergence is.

Divergence is a measure of how one distribution differs from another. It is not symmetrical in nature i.e. score of the divergence for distributions p and q would give a different score from q and p.

Here we will talk about one specific type of divergence, the Kullback-Leibler (KL) divergence.

KL divergence: It quantifies how much one probability distribution differs from another probability distribution.

The KL divergence between the two distributions p and q can be calculated using the formula:

$$D_{KL}(p||q) = \sum_{i=1}^N p(x_i) \cdot \log \frac{p(x_i)}{q(x_i)}$$

Two important points to note about KL divergence:

The lower the KL divergence value, the better the two distributions match. If two distributions perfectly match, then it is zero.

It is not symmetrical i.e.

$$D_{KL}(p||q) \neq D_{KL}(q||p)$$

The intuition behind the KL divergence score: When the probability for an event from p is large, but the probability for the same event in q is small, there is a large divergence. When the probability from p is small and the probability from q is large, there is also a large divergence, but not as large as in the former case.

To know more about KL divergence, please refer to this video [here](#).

This video talks about the intuition behind KL divergence by explaining it with a numerical example. It also provides a real-life application of this concept.

Note: In Python, we use the `rel_ent` function to calculate the KL divergence.

[← Previous](#)[Next →](#)