

KL Divergence and Entropy

Bikramjot Hanzra

July 2016

KL Divergence is a measure that can be used to compare 2 distributions i.e. how similar or dissimilar they are. In these notes we only consider discrete distribution, although there is not much change for continuous distribution. It's just that we go integration instead of summation. The KL Divergence between 2 distribution $P(x)$ and $Q(x)$ is given by –

$$D_{KL}(P||Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)} \quad (1)$$

A few quick notes on KL divergence is that –

- KL Divergence is always non-negative.
- KL Divergence is not a true distance metric as $D_{KL}(P||Q)$ is not the same as $D_{KL}(Q||P)$.

Entropy is a measure of randomness in a distribution and is given by –

$$H(x) = - \sum_x P(x) \log P(x) \quad (2)$$

Equation 1 and 2 look pretty similar. So, the question we are trying to ask here is that is there a relationship between the KL Divergence and Entropy?

The answer is yes. Entropy of a distribution can be derived from KL Divergence when we measure how far is the distribution away from the uniform distribution \mathcal{U} . So, the entropy of a distribution P can be derived using KL Divergence between the distribution P and the uniform distribution \mathcal{U} as –

$$D_{KL}(P||\mathcal{U}) = \sum_x P(x) \log \frac{P(x)}{\mathcal{U}(x)} \quad (3)$$

$$D_{KL}(P||\mathcal{U}) = \sum_x P(x) \log P(x) - \sum_x P(x) \log \mathcal{U}(x) \quad (4)$$

$$D_{KL}(P||\mathcal{U}) = -H(x) - \sum_x P(x) \log \mathcal{U}(x) \quad (5)$$

If there are m possibilities, then the uniform distribution \mathcal{U} is given as –

$$\mathcal{U}(x) = \frac{1}{m} \quad (6)$$

Substituting this in Equation 5, we get –

$$D_{KL}(P||\mathcal{U}) = -H(x) - \sum_x P(x) \log \mathcal{U}(x) \quad (7)$$

$$D_{KL}(P||\mathcal{U}) = -H(x) - \sum_x P(x) \log \frac{1}{m} \quad (8)$$

$$D_{KL}(P||\mathcal{U}) = -H(x) - \log \frac{1}{m} \sum_x P(x) \quad (9)$$

$$D_{KL}(P||\mathcal{U}) = -H(x) + \log m \quad (10)$$

$$(11)$$

In Equation 10, $\log m$ is a constant. From this equation, we can infer that the KL divergence between the distribution P and the uniform distribution \mathcal{U} is the negative entropy plus a constant term $\log m$.