KL Divergence

- It is a measure of how one probability distribution is different from the second
- It is also called relative entropy

- It is not the distance between two distribution – often misunderstood
- (Divergence is not distance)

- Jensen-Shannon divergence calculates the distance of one probability distribution from another.

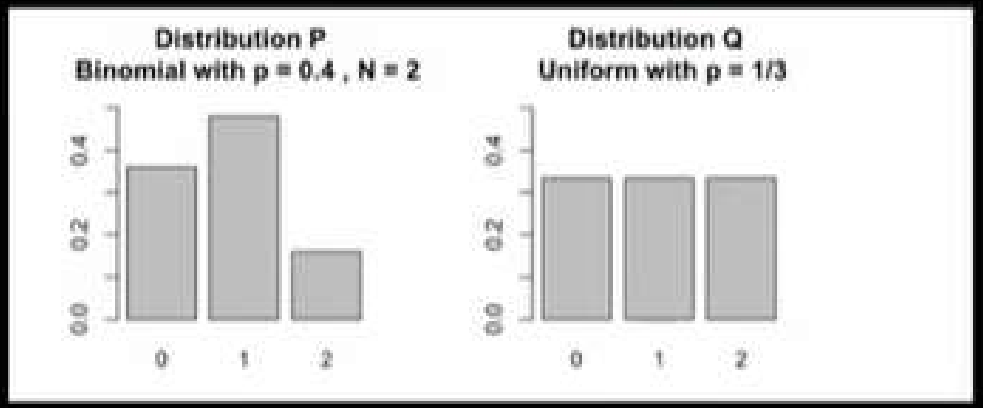
Expression

For contineous probability distributions, replace summation with integration in the formula

$$D_{KL}(P \| Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

P and Q are discrete probability distribution

Example



x	0	1	2
Distribution $P(x)$	0.36	0.48	0.16
Distribution $Q(x)$	0.333	0.333	0.333

Example

$$
\begin{aligned}
D_{\mathrm{KL}}(P \parallel Q) &= \sum_{x \in \mathcal{X}} P(x) \ln\left(\frac{P(x)}{Q(x)}\right) \\
&= 0.36 \ln\left(\frac{0.36}{0.333}\right) + 0.48 \ln\left(\frac{0.48}{0.333}\right) + 0.16 \ln\left(\frac{0.16}{0.333}\right) \\
&= 0.0852996 \\
D_{\mathrm{KL}}(Q \parallel P) &= \sum_{x \in \mathcal{X}} Q(x) \ln\left(\frac{Q(x)}{P(x)}\right) \\
&= 0.333 \ln\left(\frac{0.333}{0.36}\right) + 0.333 \ln\left(\frac{0.333}{0.48}\right) + 0.333 \ln\left(\frac{0.333}{0.16}\right) \\
&= 0.097455
\end{aligned}
$$

Interpretation

- D(P || Q) is the information gain when distribution Q is used instead of distribution P