Appendix.04 Information Theory

Definition.A.4.1. Entropy

A measure of the uncertainty of a random variable

$$\mathrm{H}[x] = -\sum_x p(x)\log p(x)$$

$(-\log p(x))$ that is amount of information quantified as # of bits required to describe the random variable.

Definition.A.4.2. Differential entropy

Let X be a random variable with a probability density function f whose support is a set $\mathcal{X}$. The differential entropy $\mathrm{H}[X]$ or $\mathrm{H}[f]$ is defined as

$$\mathrm{H}[X] = -\int_{\mathcal{X}} f(x)\log f(x)dx$$

Definition.A.4.3. Kullback-Leibler Divergence

$$\begin{aligned}\mathrm{KL}(p\|q) &= -\int p(x)\ln q(x)dx - (-\int q(x)\ln q(x)dx) \\ &= -\int p(x)\ln\{\frac{q(x)}{p(x)}\}dx\end{aligned}$$

KL divergence(= releative entropy) measures dissimilarity between two distributions.

Reference.

Deep Learning - Yosha Benjio
https://ko.wikipedia.org/wiki/정보_이론
(https://ko.wikipedia.org/wiki/%EC%A0%95%EB%B3%B4_%EC%9D%B4%EB%A1%A0)