Information theory

Siddharth Bhat

Contents

0.1 Preliminary definitions

Definition 1 *Entropy(H): The entropy of a random variable X with probability distribution $p : X \to \mathbb{R}$ is defined as:*

$$H(X) \equiv -\sum_{x \in X} p(x) \log p(x) = \mathbb{E}[-\log \circ p]$$

Definition 2 *Condtional entropy($H(X|Y)$): The conditional entropy of a random variable X with respect to another variable Y is defined as:*

$$\begin{aligned} H(X|Y) &\equiv -\sum_{y \in Y} p(y) H(X|Y = y) \\ &= \sum_{y \in Y} p(y) \sum_{x \in X} -p(x|y) \log p(x|y) \\ &= \sum_{y \in Y} \sum_{x \in X} -p(y) p(x|y) \log p(x|y) \\ &= \sum_{y \in Y} \sum_{x \in X} -p(y \wedge x) \log p(x|y) \end{aligned}$$

Definition 3 *Kullback-Leibler divergence $D(X||Y)$: The Kullback-Leibler divergence of $X \sim p$ with respect to $X' \sim q$ is:*

$$D(X||X') \equiv \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$$

Note that $D(X||X')$ is *not symmetric*.

Intuition: extra cost of encoding X if we thought the distribution were X'.

Useful extremal case to remember: Assume X' has $q(x) = 0$ for some letter $x \in X$. In this case, $D(X||X')$ would involve a term $\frac{p(x)}{0}$, which is ∞. This is intuitively sensible, since X' has no way to represent x, and hence X' is *infinitely far away from encoding* X. However, In this same case, one could have that X is able to encode all of X'.