Cheatsheet InfoTheory

Nicolas Wehrli

June 2024

1 Foundations

Definitions

Information of an outcome x

$$h(x) = -\log(p(x))$$

Cross-Entropy between p and q

$$H(p;q) = -\sum_x p(x)\log q(x)$$

Shannon Entropy

$$H(p) = H(p;p)$$

Notation
We identify outcomes x with integers $1, ..., m$ and associate probabilities $p(x) \geq 0$.
$H(\frac{1}{m})$ for $H(p)$ with $p(x) = \frac{1}{m}$ (uniform)
$H(X) = H(p) = \mathbb{E}(-\log(p(X)))$ where p is the pdf of X

Jensen's Inequality

Let f be convex and $g : [m] \to \mathbb{R}$ be an arbitrary function that assigns a value to each outcome.

$$f\left(\sum_x p(x)g(x)\right) \leq \sum_x p(x)f(g(x)), \forall p(x) \geq 0, \sum_x p(x) = 1$$

alternatively

$$f(\mathbb{E}(g(X))) \leq \mathbb{E}(f(g(X)))$$

Applying this inequality to relate Cross-Entropy and Entropy, we get the following properties.
By Jensen we have

$$H(p;q) \geq H(p)$$

Definining KL divergence or Relative Entropy as

$$D(p||q) = \sum_x p(x)\log\frac{p(x)}{q(x)}$$

we get

$$H(p;q) = H(p) + D(p||q)$$

Further investigating KL divergence, we find

$$D(p||q) \geq 0 \tag{1}$$

$$D(p||q) = 0 \iff p = q \tag{2}$$