

Cross Entropy $H(p,q)$

What if we encode X with a code optimized for a wrong distribution q ?

Expected # of bits=? $H(p,q) = E_p[-\log q(x)] = -\sum_{x \in \Omega} p(x) \log q(x)$

Intuitively, $H(p,q) \geq H(p)$, and mathematically,

$$\begin{aligned} H(p,q) - H(p) &= \sum_{x \in \Omega} p(x) \left[-\log \frac{q(x)}{p(x)} \right] \\ &\geq -\log \sum_{x \in \Omega} \left[p(x) \frac{q(x)}{p(x)} \right] = 0 \end{aligned}$$

By Jensen's inequality: $\sum_i p_i f(x_i) \geq f(\sum_i p_i x_i)$

where, f is a convex function, and $\sum_i p_i = 1$