Information and Entropy

Shannon Information: measure of surprise, or uncertainty, in event x with probability $P(x)$: $h(x) = -\log_2 P(x)$. Continuous, additive and symmetric.

Entropy: measure of disorder in random variable $X = \{x_1, x_2, ..., x_n\}$, with probability distribution $P(X)$. When we resolve disorder, we gain information.

$$H(X) = \sum_{i=1}^{n} P(x_i) \log_2 \frac{1}{P(x_i)} = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i)$$

Conditional entropy: measure of uncertainty in random variable Y given X.

$$H(Y|X=x) = -\sum_y P(y|x) \log_2 P(y|x)$$
$$H(Y|X) = \sum_x P(x) H(Y|X=x) = -\sum_x \sum_y P(x,y) \log_2 P(y|x)$$

Joint entropy: measure of uncertainty in two random variables X and Y.

$$H(X,Y) = -\sum_x \sum_y P(x,y) \log_2 P(x,y)$$

Mutual information: measure the common information between two RVs, i.e., how much information one RV conveys about another.

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) \\ &= H(X) - H(X|Y) = H(Y) - H(Y|X) \end{aligned}$$

Channel capacity: maximum mutual information achievable between input and output random variables of a channel.

$$C = \max_{p(x)} I(X;Y)$$

Probability distributions comparison and ML

Cross entropy of distributions $p(x)$ and $q(x)$:

$$H(p,q) = -\sum_x p(x) \log_2 q(x)$$

Relative entropy / Kullback-Leibler divergence of $p(x)$ from $q(x)$:

$$D_{KL}(p||q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)}$$

Relation between cross entropy and KL divergence:

$$H(p, q) = H(p) + D_{KL}(p||q)$$