1. Definitions

$$P_X(x) = \sum_y P_{(X,Y)}(x,y) \qquad \text{Marginal Dist.}$$

$$P_Y(x) = \sum_y P_{(X,Y)}(x,y) \qquad \text{Marginal Dist.}$$

$$P_{X|y}(x) = \frac{P_{(X,Y)}(x,y)}{P_Y(y)} \qquad \text{Conditional Dist.}$$

$$P_{X|Y}(x,y) = \frac{P_{(X,Y)}(x,y)}{P_Y(y)}$$

$$H[P_X(X)] = E_{x\sim P_X}[-\log P_X(x)] \qquad \text{Entropy}$$

$$H[X] = E_{x\sim P_X}[-\log P_X(x)] \qquad \text{Entropy when distribution is obvious}$$

$$H[p] = E_p[-\log p] \qquad \text{Entropy when variable is obvious}$$

$$H[X|y] = E_{x\sim P_{(X|y)}}[-\log P_{X|y}(x)] \qquad \text{Entropy of Conditional Dist.}$$

$$H[X|Y] = E_{y\sim P_Y}[H[X|y]] = E_{x,y\sim P_{(X,Y)}}[-\log P_{X|Y}(x,y)] \qquad \text{Conditional Entropy}$$

$$H[p,q] = \mathrm{E}_p[-\log q] \qquad \text{Cross Entropy}$$

$$D_{KL}(p||q) = H_{x\sim p}\left[\log\frac{p(x)}{q(x)}\right] = H[p,q] - H[p] \qquad \text{KL Divergence}$$

$$\mathrm{pmi}(x;y) = \log\frac{P_{(X,Y)}(x,y)}{P_X(x)P_Y(y)} \qquad \text{Pointwise Mutual Information}$$

$$I(X;Y) = D_{\mathrm{KL}}\left(P_{(X,Y)}\|P_X\otimes P_Y\right) = E_{x,y\sim P_{(X,Y)}}\left[\log\frac{P_{(X,Y)}(x,y)}{P_X(x)P_Y(y)}\right] = E_{x,y\sim P_{(X,Y)}}[\mathrm{pmi}(x;y)] \qquad \text{Mutual Information}$$

2. Identities

$$H[p,q] \geq H[p] \qquad \text{Gibb's Inequality}$$

$$D_{KL}(p||q) = H[p,q] - H[p] \geq 0$$

$$\mathrm{H}(X,Y) = \mathrm{H}(X|Y) + \mathrm{H}(Y) = \mathrm{H}(Y|X) + \mathrm{H}(X) \qquad \text{Chain Rule}$$

$$\begin{aligned}\mathrm{I}(X;Y) &\equiv \mathrm{H}(X) - \mathrm{H}(X\mid Y) = E_{x,y\sim P_{(X,Y)}}[-\log P_X(x)] - E_{x,y\sim P_{(X,Y)}}[-\log\frac{P_{(X,Y)}(x,y)}{P_Y(y)}]\\ &\equiv \mathrm{H}(Y) - \mathrm{H}(Y\mid X)\\ &\equiv \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X,Y)\\ &\equiv \mathrm{H}(X,Y) - \mathrm{H}(X\mid Y) - \mathrm{H}(Y\mid X)\end{aligned}$$

3. Useful Inequalities

1. Log-sum inequality: $\sum_{i=1}^n a_i \log\frac{a_i}{b_i} \geq a\log\frac{a}{b}$
    1. $a = \sum_i a_i$, $b = \sum_i b_i$
    2. $a_i \geq 0$, $b_i \geq 0$