Define entropy and list some properties

Cover and Thomas

Define conditional entropy and list some properties

Cover and Thomas

Define mutual information and list some properties

Cover and Thomas

Definition. *For a finite-valued random variable X,* $H(X) = -\sum_i p(x_i) \log p(x_i)$.

Properties.

1. *Entropy is non-negative $H(X) \geq 0$ (for finite-ranged rvs).*
2. *Entropy is bounded by the log of the size of the range $H(X) \leq \log |\mathcal{X}|$.*
3. *It is concave: $H(\lambda p_1(x) + (1-\lambda)p_2(x)) \geq \lambda H(p_1(x)) + (1-\lambda)H(p_2(x))$.*

Definition. *For two random variable X and Y, the conditional entropy is*

$$H[X|Y] = \sum_y p(y) H[X|Y = y]$$

where

$$H[X|Y = y] = \sum_x p(x|y) \log p(x|y)$$

Properties.

-

Definition. *For two finite random variables X and Y, with joint density $p(x, y)$ and marginal densities $p(x), q(y)$ respectively, the mutual information is*

$$I(X, Y) = KL(p(x, y) \| p(x)q(y)) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)q(y)}$$

Properties.

-

Define the Kullback-Leibler divergence and list some properties

Cover and Thomas

State the data processing inequality and give a use case

Cover and Thomas

State Fano's inequality and give a use case

Cover and Thomas

Definition. *For two finite random variables X and Y, with densities $p(x), q(x)$ respectively, the Kullback-Leibler divergence is*

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

Properties.

-

Definition. *The data processing inequality says that for three random variables X, Y, Z satisfying the Markov chain condition $X \to Y \to Z$ have $I(X,Y) \geq I(X,Z)$.*

Example.

Definition. *Fano's inequality states that for any two finite random variables $X, \hat{X}$, taking values in $\mathcal{X}$, we have*

$$h(p) + p \log \mathcal{X} \leq H(X|\hat{X})$$

where $p = \Pr\left(X \neq \hat{X}\right)$.

Example.

Cover and Thomas