

Information theoretic learning notes

By Manuel Pineda

Shannon Entropy and information unit.

Self information:

$$\mathbf{I}(X) = \log \frac{1}{\mathbf{P}(X)}$$

Shannon entropy of a random variable X (expected value over the self information of X):

$$\mathbf{H}(X) = \mathbf{E}[\mathbf{I}(X)] = \mathbf{E}[-\log \mathbf{P}(X)]$$

Conditional entropy:

$$\mathbf{H}(X \mid Y) = -\sum_{i,j} p(x_i, y_j) * \log \frac{p(x_i, y_j)}{p(y_i)}$$

Properties of shannon entropy:

- Continuity
- Symmetry: $\mathbf{H}(p_1, p_2, \dots, p_n) = \mathbf{H}(p_2, p_1, \dots, p_n)$
- Maximum:
- Additivity