Lecture Notes

on

Information Theory and Coding

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

July 2020

(Be safe and stay at home)

Mutual Information

- The mutual information of two random variables X and Y is defined as the relative entropy between the joint probability density $p_{XY}(x,y)$ and the product of the marginals $p_X(x)$ and $p_Y(y)$

$$
\begin{aligned}
I(X;Y) &= D(p_{XY}(x,y)||p_X(x)p_Y(y)) \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{XY}(x,y)\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}
\end{aligned}
$$

Mutual Information: Relations with Entropy

- Reducing uncertainty of X due to the knowledge of Y:

$$I(X;Y) = H(X) - H(X|Y)$$

- Symmetry of the relation above: $I(X;Y) = H(Y) - H(Y|X)$
- Sum of entropies:

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

- "Self" Mutual Information:

$$I(X;X) = H(X) - H(X|X) = H(X)$$

Mutual Information: Other Relations

- Conditional Mutual Information:

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$$

- Chain Rule for Mutual Information

$$I(X_1, X_2, \ldots, X_M; Y) = \sum_{j=1}^{M} I(X_j; Y|X_1, \ldots, X_{j-1})$$

Convex and Concave Functions

- A function $f(\cdot)$ is convex over ain interval (a, b) if for every $x_1, x_2 \in [a, b]$ and $0 \leq \lambda \leq 1$, if :

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$$

- A function $f(\cdot)$ is convex over an interval (a, b) if its second derivative is non-negative over that interval (a, b).
- A function $f(\cdot)$ is concave if $-f(\cdot)$ is convex.
- Examples of convex functions: x^2, $|x|$, e^x, $x \log x, x \geq 0$.
- Examples of concave functions: $\log x$ and $\sqrt{x}$, for $x \geq 0$.

Jensen's Inequality

- If $f(\cdot)$ is a convex function and X is a random variable

$$E[f(X)] \geq f(E[X])$$

- Used to show that relative entropy and mutual information are greater than zero.
- Used also to show that $H(X) \leq \log |\mathcal{X}|$.

Log-Sum Inequality

- For n positive numbers $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots b_n$

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i\right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}$$

  with equality if and only if $a_i/b_i = c$.
- This inequality is used to prove the convexity of the relative entropy and the concavity of the entropy.
- Convexity/Concavity of mutual information

Data Processing Inequality

- Random variables X, Y, Z are said to form a Markov chain in that order $X \rightarrow Y \rightarrow Z$, if the conditional distribution of Z depends only on Y and is onditionally independent of X.

$$p_{XYZ}(x, y, z) = p_X(x)p_{Y|X=x}(y)p_{Z|Y=y}(y)$$

- If $X \rightarrow Y \rightarrow Z$, then

$$I(X; Y) \geq I(X; Z)$$

- Let $Z = g(Y)$, $X \rightarrow Y \rightarrow g(Y)$, then $I(X; Y) \geq I(X; g(Y))$

Fano's Inequality

- Suppose we know a random variable Y and we wish to guess the value of a correlated random variable X.
- Fano's inequality relates the probability of error in guessing X from Y to its conditional entropy $H(X|Y)$.
- Let $\hat{X} = g(Y)$, if $P_e = P(\hat{X} \neq X)$, then

$$H(P_e) + P_e \log(|\mathcal{X}| - 1) \geq H(X|Y)$$

where $H(P_e)$ is the binary entropy function evaluated at P_e.