Entropy and Mutual Information

M.Tech CS
Information and Coding Theory

Before Mid Semester

1 Entropy

Definition: 1. *Entropy is the minimum descriptive complexity of a random variable. In other word, it is measure of the uncertainty of a random variable.*

Formally,
If $\boldsymbol{X}$ be a discrete random variable over alphabet χ and probabilty mass function is p(x)=Pr{ $X = x$ }= x $x \in \chi$.
The Entropy *H(χ)* of a discrete random variable $\boldsymbol{X}$ is defined by

$$H(X) = -\sum_{x \in \chi} p(x) log_2 p(x) \tag{1}$$

Entropy is expressed in bit. When entropy is measured in natural base (ln) it is called *nats*.
Some useful properties

1. $H(X) \geq 0$
2. $H_b(X) = (\log_b a H_a(X)$.
3. $H(X) = -E \log p(x) = E \frac{1}{\log p(X)}$ where E($\bullet$) is expection of given function.

1.1 Joint Entropy and Conditional Entropy

Definition: 2. *The **joint entropy** $H(X,Y)$ of a pair of discrete random variables (X,Y) with joint distribution $p(x,y)$ is defined as*

$$H(X,Y) = -\sum_{x \in \chi} \sum_{y \in \gamma} p(x,y) log_2 p(x,y) \tag{2}$$

$$H(X,Y) = -E log_2 p(X,Y) \tag{3}$$

Definition: 3. *if (X,Y) $\sim$ p(x,y), then **conditional entropy** $H(Y|X)$ is defined as*

$$
\begin{aligned}
H(Y|X) &= \sum_{x \in \chi} p(x) H(Y|X=x) \\
&= -\sum_{x \in \chi} p(x) \sum_{y \in \gamma} p(y|x) \log_2 p(y|x) \\
&= -\sum_{x \in \chi} \sum y \in \gamma p(x,y) \log_2 p(y|x) \\
&= -E \log_2 p(Y|X)
\end{aligned}
$$

Theorem: 1. *Chain Rule:*

$$
H(X,Y) = H(X) + H(Y|X)
$$

Proof.

$$
\begin{aligned}
H(X,Y) &= -\sum_{x \in \chi, y \in \gamma} p(x,y) \log_2 p(x,y) \\
&= -\sum_{x \in \chi} \sum_{y \in \gamma} p(x,y) \log_2 p(x) p(y|x) \\
&= -\sum_{x \in \chi} \sum_{y \in \gamma} p(x,y) \log_2 p(x) - \sum_{x \in \chi} \sum_{y \in \gamma} p(x,y) \log_2 p(y|x) \\
&= -\sum_{x \in \chi} p(x) \log_2 p(x) - \sum_{x \in \chi} \sum_{y \in \gamma} p(x,y) \log_2 p(y|x) \\
&= H(X) + H(Y|X)
\end{aligned}
$$

□

1.2 Relative Entropy

Definition: 4. ***Relative Entropy*** *or* ***Kullback-Liebler distance*** *between two probabilty mass functions $p(x)$ and $q(x)$ over alphabet χ*

$$
D(p(x)||q(x)) = \sum_{x \in \chi} p(x) \log_2 \frac{p(x)}{q(x)} \tag{4}
$$

$$
D(p(x)||q(x)) = E_{p(x,y)} \log_2 \frac{p(X)}{q(X)} \tag{5}
$$

When p(x) = q(x) for $x \in \chi$
$D(p(x)||q(x)) = 0$

1.3 Mutual Information

Definition: 5. *Two random variables X and Y with a joint probability mass function $p(x,y)$ and marginal probability mass functions $p(x)$ and $p(y)$, then* ***Mutual Information*** *is defined as* *relative entropy of joint probability mass function* $p(x, y)$ *and product of marginal probability* $p(x)$ *and* $p(y)$

$$I(X;Y) = \sum_{x \in \chi} \sum_{y \in \gamma} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \tag{6}$$

$$= D(p(x, y)||(p(x)p(y))$$

$$= E_{p(x,y)} \log_2 \frac{p(X, Y)}{p(X)p(Y)}$$