$$\text{Mutual Information} = \sum_{y=Y} \sum_{x=X} p(x,y) \cdot \log \frac{p(x,y)}{p(x)\,p(y)}$$

$$\text{Entropy} = H(Y) = -\sum_{y \in Y} p(y) \cdot \log(p(y))$$

$$\text{Conditional Entropy} = H(Y|X) = \sum_{x \in X} p(x) \cdot H(Y|X=x)$$

Show mathematically that Mutual Information = Information Gain, where information gain = $\boxed{H(Y) - H(Y|X)}$.

Solution

We know that mutual information $I(X;Y)$ is mathematically defined as

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

This helps us quantify how much uncertainity the realization of a random variable X has if the outcome of another random variable Y is known.

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

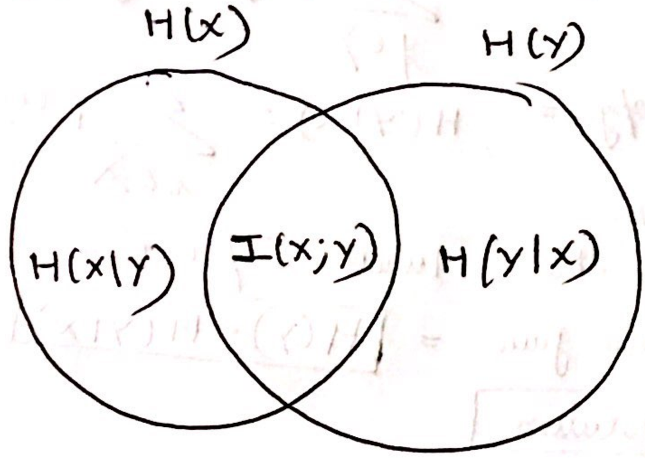
$$= \sum_{x \in X} P_X(x) \cdot \log_2 \frac{1}{P_X(x)} + \sum_{y \in Y} P_Y(y) \cdot \log_2 \frac{1}{P_Y(y)}$$

$$- \sum_{x \in X} \sum_{y \in Y} P(x,y) \cdot \log_2 \frac{1}{P_{XY}(x,y)}$$

$$= \sum_{x \in X} \sum_{y \in Y} P_{XY}(x,y) \cdot \log_2 \frac{P_{XY}(x,y)}{P_X(x)\,P_Y(y)}$$

Information gain : $H(Y) - H(Y|X)$ mathematically

Looking at the Venn diagram below



Conditional entropy is the remaining parts to generate if we generate mutual information first.

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

$$= H(Y) - [H(X) + H(Y) - H(XY)]$$

$$= -H(X) + H(XY) = H(XY) - H(X).$$

$$= \sum_{x \in X} \sum_{y \in Y} P_{XY}(x,y) \cdot \log_2 \frac{1}{P_{XY}(x,y)} - \sum_{x \in X} P_X(x) \log_2 \frac{1}{P_X(x)}$$

Now Information gain $= H(Y) - H(Y|X)$

$$= H(Y) - [H(XY) - H(X)]$$

$$= H(Y) + H(X) - H(XY)$$

$$= \sum_{x \in X} \sum_{y \in Y} P_{XY}(x,y) \log_2 \frac{P_{XY}(x,y)}{P_X(x) P_Y(y)}$$

= Mutual Information.

Thus we see mutual information (or the "information gain" reduces uncertainity measures for X when Y is known