

## Code Assignment 3

1. Implement the decision tree using Python based on information gain for splitting nodes (features). The data and information gain formula is provided below:

The formula for Entropy is shown below:

$$E(S) = -p_{(+)} \log p_{(+)} - p_{(-)} \log p_{(-)}$$

Here,

- $p_{+}$  is the probability of positive class
- $p_{-}$  is the probability of negative class
- $S$  is the subset of the training example

### Information Gain

Information gain measures the reduction of uncertainty given some feature and it is also a deciding factor for which attribute should be selected as a decision node or root node.

$$\text{Information Gain} = E(Y) - E(Y|X)$$

## Training Dataset

The class label attribute, *buys\_computer*, has two distinct values.

Thus there are two distinct classes. ( $m=2$ )

Class C1 corresponds to *yes* and class C2 corresponds to *no*.

There are 9 samples of class *yes* and 5 samples of class *no*.

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no