Answer according to Question No 1

ID: 18030 1280
Md Labib Aretin

Information gain is the reduction in entropy or surprise by transforming a data set and is used in training decision trees.

$H(Y|x)$ means;

- I know x
- It tells the uncertainty in Y after knowing x

$H(Y|x) = H(Y) \rightarrow x$ and Y are statistically independent

If x and Y are dependented, and I know x, for this if Y's uncertainty is decreased, then it is called information gain

$IG(Y, x) = H(Y) - H(Y|x) = 0$

where,

$x = 0$, when $H(Y|x) = H(Y) : Y \perp x$

$x = +ve$, when $H(Y) > H(Y|x)$

$x = -ve$ when $H(Y) < H(Y|x)$

This is where the math starts;-

$P(\text{Play Golf} = Y) = 9/14 = 0.64$

$P(\text{Play Golf} = N) = 5/14 = 1 - 0.64 = 0.36$

$$H(\text{Play Golf}) = -\sum_{a \in A} P(a) \log_2 P(a)$$

$$= -\left[0.64 \log_2(0.64) + 0.36 \log_2(0.36)\right]$$

$$= -\left[-0.64 \times 0.644 + (-1.474 \times 0.36)\right]$$

$$= 0.9428$$

$$P(\text{Outlook} = \text{Rainy}) = \frac{5}{14} = 0.35 \qquad P(\text{Outlook} = \text{Sunny}) = \frac{5}{14} = 0.35$$

$$P(\text{Outlook} = \text{Overcast}) = \frac{4}{14} = 0.30$$

$$H(\text{Play Golf} \mid \text{Outlook} = \text{Rainy}) = -\sum_{a \in A} \left[P(\text{Play Golf} \mid \text{Outlook} = 'R') * \log_2 P(\text{Play Golf} \mid \text{Outlook} = 'R') \right]$$

$$= -\sum_{a \in A} \left[P(\text{Play Golf} = Y \mid \text{Outlook} = 'R') * \log_2 P(\text{Play Golf} = Y \mid \text{Outlook} = 'R') + P(\text{Play Golf} = N \mid \text{Outlook} = 'R') * \log_2 P(\text{Play Golf} = N \mid \text{Outlook} = 'R') \right]$$

$$P(P = Y \mid O = r) = \frac{P_r[P = Y, O = r)}{P_r(O = r)}$$

$$= \frac{\frac{2}{14}}{0.35} = \frac{1}{20} = \frac{2}{40}$$

$$P_r(P = N \mid O = r) = \frac{P(P = N \mid O = r)}{P(O = r)} = \frac{\frac{3}{14}}{0.35} = \frac{3}{40}$$

$$P_r(P = Y \mid O = s) = \frac{P(P = Y \mid O = s)}{P(O = s)} = \frac{3/14}{0.35} = \frac{3}{40}$$

$$P_r(P = N \mid O = s) = \frac{P(P = N \mid O = s)}{P(O = s)} = \frac{2/14}{0.35} = \frac{2}{40}$$

$$P_r(P = Y \mid O = o) = P(P = Y \mid O = o) = \frac{4/14}{0.3} = \frac{10}{21}$$

$$H(\text{Play Golf} \mid \text{Outlook} = r) = -\left[\frac{2}{40}\log_2\frac{2}{40} + \frac{3}{40}\log_2\frac{3}{40}\right]$$

$$= -\left[\frac{2}{40} \times (-4.32) + \frac{3}{40} \times (-3.74)\right]$$

$$= -(-0.216 - 0.2801)$$

$$= 0.497$$

$$H(\text{Play Golf} \mid \text{Outlook} = s) = -\left[\frac{3}{40}\log_2\frac{3}{40} + \frac{2}{40}\log_2\frac{2}{40}\right]$$

$$= \left[-3.74 \times \frac{3}{40} + \frac{2}{40} \times (-4.32)\right]$$

$$= 0.497$$

$$H(\text{Play Golf} \mid \text{Outlook} = o) = \left[\frac{20}{21}\log_2\left(\frac{20}{21}\right)\right]$$

$$= 0.070$$

This is where (V) started.
conditional entropy

$$H(\text{Play Golf} \mid \text{Outlook}) = \sum_{a \in A} P(a)\, H(\text{Play Golf} \mid \text{Outlook} = a)$$

$$= \frac{5}{14} \times 0.497 + \frac{5}{14} \times 0.497 + \frac{4}{14} \times 0.070$$

$$= 0.1775 + 0.1775 + 0.02$$

$$= 0.375$$

$$IG(\text{Play Golf} \mid \text{Outlook}) = 0.9428 - 0.375$$

$$= 0.5678$$

0411803012BD – (4)

Answer according to Question No 2

AI

AI (Artificial Intelligence is defined as acquisition of knowledge and ability to apply knowledge

The aim is to simulate natural intelligence to solve tough issues

ML

It is practice of getting machines to make decisions without being programmed

ML is a subset of AI and data science

DL

In deep learning, it implements Artificial Neural Network (ANN) which has multiple layers between its input and output layers

DS

Data science is a broad field that spans to collection, management and analysis of large amounts of data with a wide range of applications.

It integrates all terms above and more to summarize insights from data and make predictions from large data sets.

04190301280 - (C)

Answer according to Question No 3

Cross validation is a resampling procedure used to evaluate machine learning models on limited data sample and it can divide data into 2 segments :

- ✓ one can used to learn or train a model
- ✓ one o used to validate the model.

* Overfitting is an error that occurs in data modelling as result of a particular function aligning to closely to a minal set of data points.

* ~~Prediction~~ pro When a model has been compromised by overfitting, the model may loose its value as a predictive tool for investing

* Overfitting is more frequent problem that underfitting and typically occurs as result of trying to avoid floating.

We can remove the problem by :-

1. model must be linear $y = ax + b$ (too simple)
2. model can be quadratic $y = ax^2 + bx + c$

$= ax_1 + bx_2 + cx_0$

(good)

* To prevent it from cross_validation, in which data being used for training model is chopped into folds and model is used for each fold.