Ans. Features = outlook, climate, wind and humidity.
To determine the splits on each level of decision tree,

$$E(I) = -\left(\frac{5}{14}\log\left(\frac{5}{14}\right) + \frac{9}{14}\log\left(\frac{9}{14}\right)\right)$$

$$= 0.943$$

Info gain (I, outlook)

$$= 0.94 - \left[\frac{5}{14}\left(-\frac{3}{5}\log\frac{3}{5} - \frac{2}{5}\log\frac{2}{5}\right) + \frac{4}{14}(0)\right.$$

$$\left. + \frac{5}{14}\left[-\frac{3}{5}\log\frac{3}{5} - \frac{2}{5}\log\frac{2}{5}\right]\right]$$

$$= 0.24$$

Infogain (I, Climate)

$$= 0.94 - \left[\frac{4}{14}\left[-\frac{2}{2}\log 1/2 + \frac{6}{14}\left(-\frac{2}{6}\log\frac{2}{6} - \frac{4}{6}\log\frac{4}{6}\right)\right.\right.$$

$$\left. + \frac{4}{14}\left(-\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4}\right)\right]$$

$$= 0.02$$

Infogain (I, Wind)

$$= 0.94 - \left[\frac{8}{14}\left[-\frac{2}{8}\log\frac{2}{8} - \frac{6}{8}\log\frac{6}{8}\right] + \frac{6}{14}\left(-\frac{2}{2}\log\frac{1}{2}\right)\right]$$

$$= 0.04$$

Infogain(S, humidity)

$$= 0.94 - \left[\frac{7}{4}\left(-\frac{4}{7}\log\frac{4}{7} - \frac{3}{4}\log\frac{3}{7}\right) + \frac{7}{14}\left(-\frac{6}{7}\log\frac{6}{7} - \frac{1}{7}\log\frac{1}{7}\right)\right]$$

$= 0.15$

Outlook

sunny → $E(2,3)$

overcast → Yes, $E(4,0) = 0$

Rain → $E(3,2)$

$E(2,3) = -\frac{3}{5}\log\frac{3}{5} - \frac{2}{5}\log\frac{2}{5}$

$= 0.97$

$E(3,2) = -\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\log\frac{3}{5}$

$= 0.97$

features available after sunny = wind, climate, humidity

~~E(0.97)~~

~~E(~~

$\text{Infogain}(S, H) = 0.97 - \left[\frac{3}{5}(0) + \frac{2}{5}(0)\right] = 0.97$

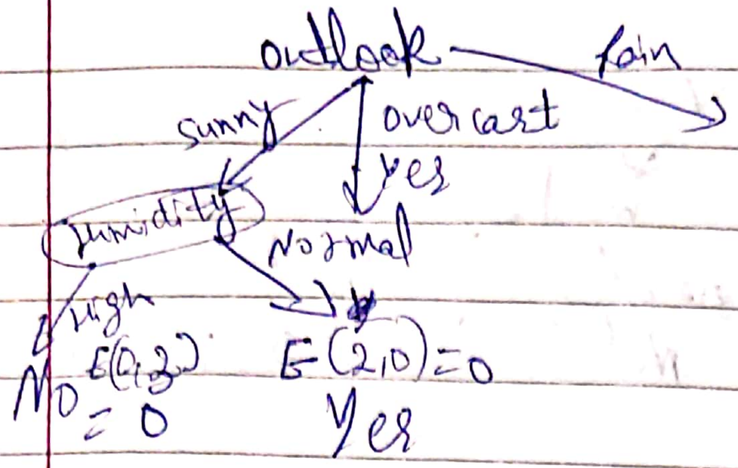
$\text{Infogain}(S, C) = 0.97 - \left[\frac{2}{5}\left(-1\log\frac{1}{2}\right)\right]$

$= 0.571$

$\text{Infogain}(S, W) = 0.971 - \left[\frac{3}{5}\left(-\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}\right) + \frac{2}{5}\left(-1\log\frac{1}{2}\right)\right]$

$= 0.02$

Scanned with CamScanner




For Rain,

$$\text{Infogain}(R,C) = 0.97 - \left[\frac{3}{5}\left(-\frac{3}{5}\log\frac{3}{5} - \frac{2}{5}\log\frac{2}{5}\right) + \frac{2}{5}\left(-\log\frac{1}{2}\right)\right]$$

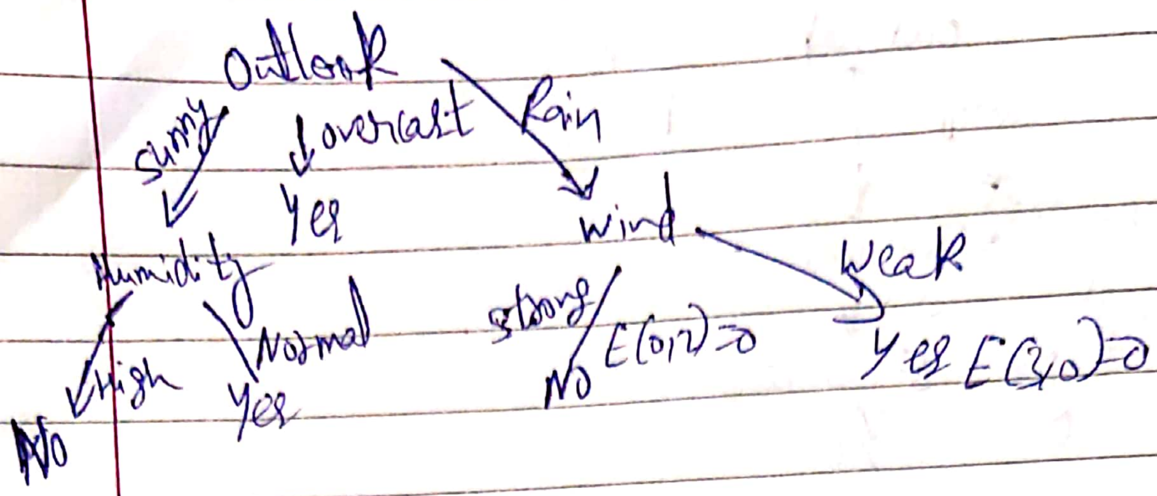
$$= 0.02$$

$$\text{Infogain}(R,H) = 0.97 - \left[\frac{3}{5}\left(-\frac{3}{5}\log\frac{3}{5} - \frac{2}{5}\log\frac{2}{5}\right) + \frac{2}{5}\left(-1\log\frac{1}{2}\right)\right]$$

$$= 0.02$$

$$\text{Infogain}(R,W) = 0.97 - \left[\frac{3}{5}\times 0 + \frac{2}{5}\times 0\right]$$

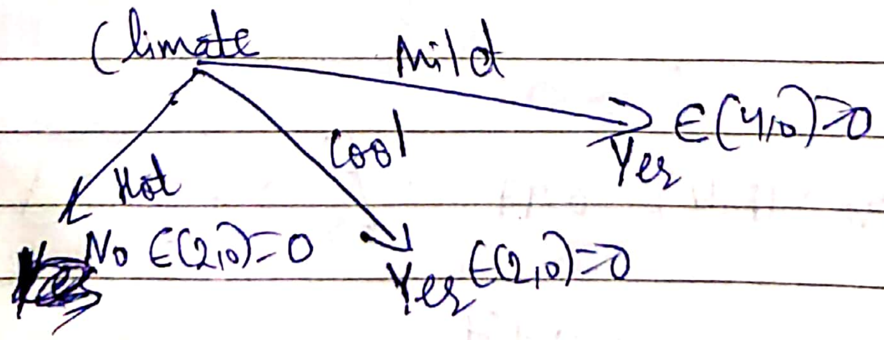
$$= 0.97$$




~~Every label is~~
~~Every label~~
Every lable is clearly defined and distinguished by this tree, so accuracy will be $\frac{14}{14} = 1$.

Q.b) Yes,

Let the training set be
< $D_1, D_2,$ ~~$D_3$~~, $D_5$, ~~$D_6$~~, ~~$D_7$~~, $D_{10}, D_{11}, D_{12}, D_{14}$




~~Case~~ Here, our climate attribute directly predicts play match.

(5)

c). for train set ($D_1$ – $D_7$)

$$E(T) = -\frac{3}{7}\log\frac{3}{7} - \frac{4}{7}\log\frac{4}{7}$$

$$= 0.98$$

$$\text{Infogain}(T, O) = 0.98 - \left[\frac{3}{7}\left(-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}\right)\right]$$

$$= 0.59$$

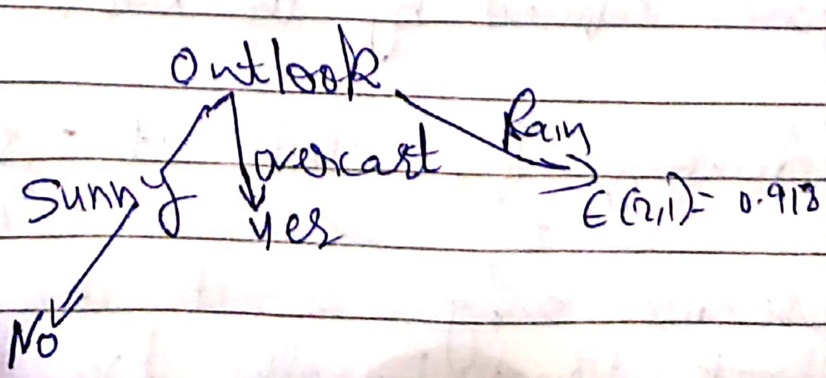
$$\text{Infogain}(T, C) = 0.98 - \left[\frac{3}{7}\left(-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}\right) + \frac{3}{7}\left(-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}\right)\right]$$

$$= 0.29$$

$$\text{Infogain}(T, W) = 0.98 - \left[\frac{4}{7}\left(-1\log\frac{1}{2}\right) + \frac{3}{7}\left(-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}\right)\right]$$

$$= 0.13$$

Tree



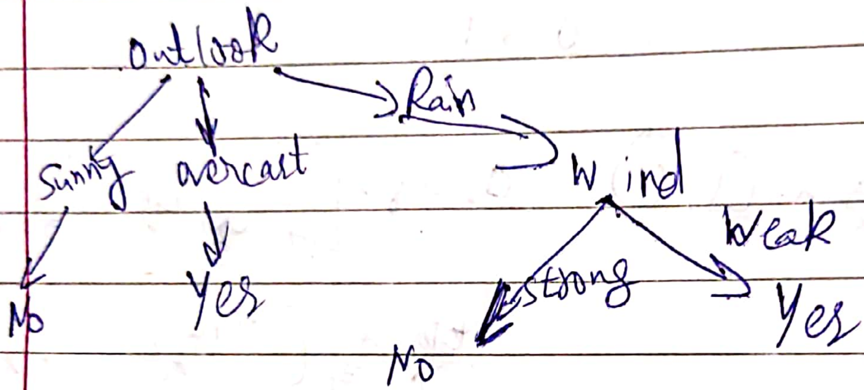
For ~~Rain~~ Rain split,

$$\text{Infogain}(R, C) = 0.91 - \left[\frac{2}{3}\left(-1 \log \frac{1}{2}\right)\right]$$

$$= 2.25$$

$$\text{Infogain}(R, W) = 0.91 - \left[\frac{2}{3} \times 0 - \frac{1}{3}(0)\right]$$

$$= 0.91$$




~~All the trainin~~

Accuracy for train ~~set 0.1~~ set = 1 because all the instances are defined by the tree.

Accuracy for test set $= \frac{5}{7} = 0.711$

because ~~only~~ 2 sets are not correctly ~~fixed~~ defined by the tree because of overfitting.

d). Pruning strategy ~~to~~ to avoid overfitting →
→ Max depth can be taken, so that it can prevent overfitting ~~a~~ of the model.

The more the number of ~~leaf~~ minimum samples required to be in leaf node, more is possibility of overfitting. So, it should be dealt carefully.

Q7.
a). Logistic regression treats all features to be uncorrelated, so it can ~~be~~ ~~when~~ cause problems where features are highly correlated with each other. But in case of DT, it establish relationship b/w features by making ~~set~~ tree of feature by branching ~~so~~, ~~So~~ which is ~~a~~ the biggest advantage of DT over logistic regression.

b). DT ~~can~~ can be overfitted when it is trained ~~and~~ under high depth and no. of leaf nodes whereas LR has low variance and is less prone to overfit.

c). Yes,
as our ~~vectors~~ are linearly
seperable , ~~vect~~
we can do our classification.
We need to split data ~~for~~ by $x_1$
only and set threshold ~~at~~ for ~~at~~
$x_2$ after $x_1$.
For n values,
depth = ~~O(log~~ n) O(log(n))

d). ~~In this~~ Yes,
In this case, feature are independent.
We can build a tree by splitting $x_1$,
then we can start splitting based on
$x_2$ at the nodes of $x_1$.
Total depth = 2 log(n)

Q6. $w_1$ = Tough, $w_2$ = course, $w_3$ = ?
$w_4$ = course

~~P(w3~~
~~Since, it is~~ markov chains are ~~used~~

Since first order markov model is used,
$w_i$ depends on $w_{i-1}$ and $w_{i+1}$ and is independent of others.

~~So~~
~~$P(w_3 / w_1, w_2, w_3)$~~ $= P(w_3 / w_2, w_4)$

$$P(w_3 / w_2, w_4) = \frac{P(w_3 / w_2)\, P(w_4 / w_2, w_3)}{P(w_3 / w_2)\, P(w_4 / w_3) + P(w_2 / w_3)}$$

$+ P(w$

$w_3$ can have values tough or course, so,

$P(w_3 = \text{tough} / w_2, w_1)$

$$= \frac{P(\text{tough} / \text{course})\, P(\text{course} / \text{tough})}{P(\text{tough} / \text{course})\, P(\text{course} / \text{tough}) + P(\text{course} / \text{course})\, P(\text{course} / \text{course})}$$

Page No.
Date

(11)

$$= \frac{0.5 \times 0.3}{0.5 \times 0.5 + 0.28}$$

$$= 0.375$$

Ans.

P (w3 = course / $w_2$, w4)

$= 1 - 0.375$

$= 0.625$

Date

(11)

Q8. $P(Y=1/X)$

$$= \frac{P(Y=1)\,P(X/Y=1)}{P(Y=1)\,P(X/Y=1) + P(Y=0)\,P(X/Y=0)}$$

$$= \frac{1}{1 + \frac{P(Y=0)\,P(X/Y=0)}{P(Y=1)\,P(X/Y=1)}}$$

$$= \frac{1}{1 + e^{\ln\left[\frac{P(Y=0)\,P(X/Y=0)}{P(Y=1)\,P(X/Y=1)}\right]}}$$

$$= \frac{1}{1 + e^{\left(\ln\frac{P(Y=0)}{P(Y=1)} + \sum_{i=1}^{n}\ln\left(\frac{P(X_i/Y=0)}{P(X_i/Y=1)}\right)\right)}}$$

$$= \frac{1}{1 + e^{\ln\left(\frac{1-\pi}{\pi}\right) + \sum_{i=1}^{n}\ln\left(\frac{P(X_i/Y=0)}{P(X_i/Y=1)}\right)}} \quad -(A).$$

(12)

Date

$\theta_{i1} = P(x_i = 1 / Y = 1)$

$\theta_{i0} = P(x_i = 1 / Y = 0)$

$1 - \theta_{i1} = P(x_i = 0 / Y = 1)$
$1 - \theta_{i0} = P(x_i = 0 / Y = 0)$

~~Using these eqn~~,
Since, they are boolean,

$P(x_i / Y = 1) = \theta_{i1}^{x_i} (1 - \theta_{i1})^{1 - x_i}$
$P(x_i / Y = 0) = \theta_{i0}^{x_i} (1 - \theta_{i0})^{1 - x_i}$

Substituting in eq (1),

$P(Y = 1 / X)$

$$= \frac{1}{1 + e^{\ln\left(\frac{1-\pi}{\pi}\right) + \sum_{i=1}^{n} \ln \left( \frac{\theta_{i0}^{x_i} (1 - \theta_{i0})^{1 - x_i}}{\theta_{i1}^{x_i} (1 - \theta_{i1})^{1 - x_i}} \right)}}$$

$$= \frac{1}{1 + e^{\left[\ln\left(\frac{1-\pi}{\pi}\right) + \sum_{i=1}^{n} \ln\left[\left(\frac{\theta_{i0}^{x_i}(1-\theta_{i1})^{x_i}}{\theta_{i1}^{x_i}(1-\theta_{i0})^{x_i}}\right)\left(\frac{1-\theta_{i0}}{1-\theta_{i1}}\right)\right]\right]}}$$

$$= \frac{1}{1 + e^{\left[\ln\left(\frac{1-\pi}{\pi}\right) + \sum_{i=1}^{n}\left[\ln\left(\frac{\theta_{i0}(1-\theta_{i1})}{\theta_{i1}(1-\theta_{i0})}\right)\cdot x_i + \ln\left(\frac{1-\theta_{i0}}{1-\theta_{i1}}\right)\right]\right]}}$$

$$= \frac{1}{1 + \exp\left(\ln\left(\frac{1-\pi}{\pi}\right) + \sum_{i=1}^{n} \ln\left(\frac{1-\theta_{i0}}{1-\theta_{i1}}\right) + \sum_{i=1}^{n} x_i \ln\left(\frac{\theta_{i0}(1-\theta_{i1})}{\theta_{i1}(1-\theta_{i0})}\right)\right)}$$

Scanned with CamScanner

Date

$P(Y=1/X)$

$$= \frac{1}{1+\exp\left[\ln\left(\frac{1-\pi}{\pi}\right) + \sum_{i=1}^{n} \ln\left(\frac{1-\theta_{i0}}{1-\theta_{i1}}\right) + \sum_{i=1}^{n} X_i \ln\left(\frac{\theta_{i0}(1-\theta_{i1})}{\theta_{i1}(1-\theta_{i0})}\right)\right]}$$

which is of the form

$$= \frac{1}{1+\exp\left(w_0 + \sum_{i=1}^{n} X_i w_i\right)}$$

Hence Proved.