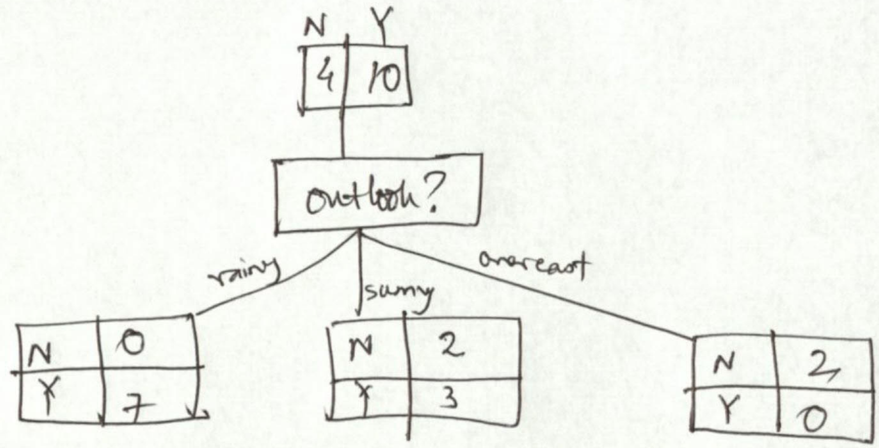
Ans to 1(a)

Gini

The goal is to maximize gain i.e. minimize gini for a selection of split

Step-1:-

For attribute outlook, multi-way split has the lowest weighted gini (0.1714). The gini index for split combinations {sunny, overcast + rainy}, {sunny + overcast, rainy}, and {overcast, sunny + rainy} are respectively 0.3936, 0.2448, 0.2381



Step-2:-

The splits for humidity has the smallest gini index; so the tree becomes:

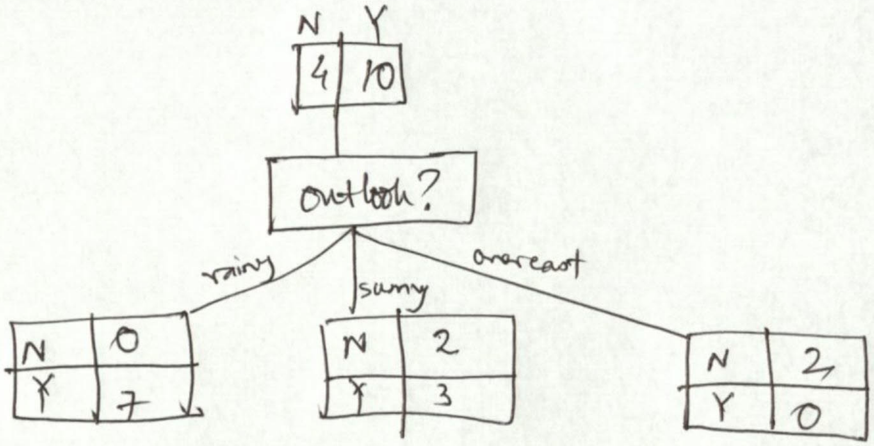
Ans to 1(a) –

Gini

The goal is to maximize gain i.e. minimize gini for a selection of split

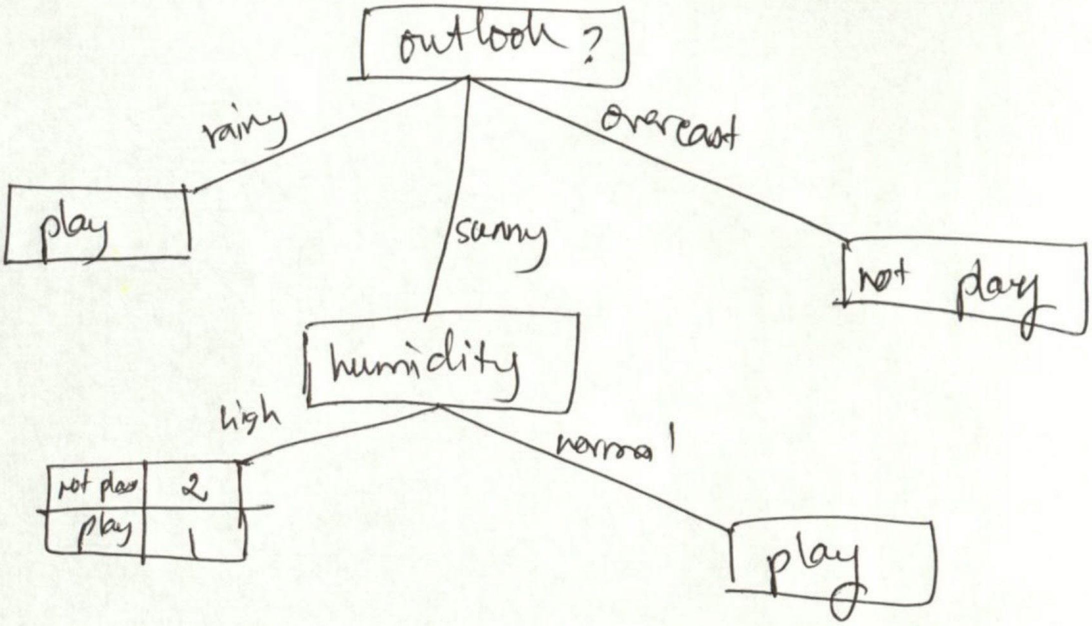
Step-1: –

For attribute outlook, multi-way split has the lowest weighted gini (0.1714). The gini index for split combinations {sunny, overcast + rainy}, {sunny + overcast, rainy}, and {overcast, sunny + rainy} are respectively 0.3936, 0.2448, 0.2381



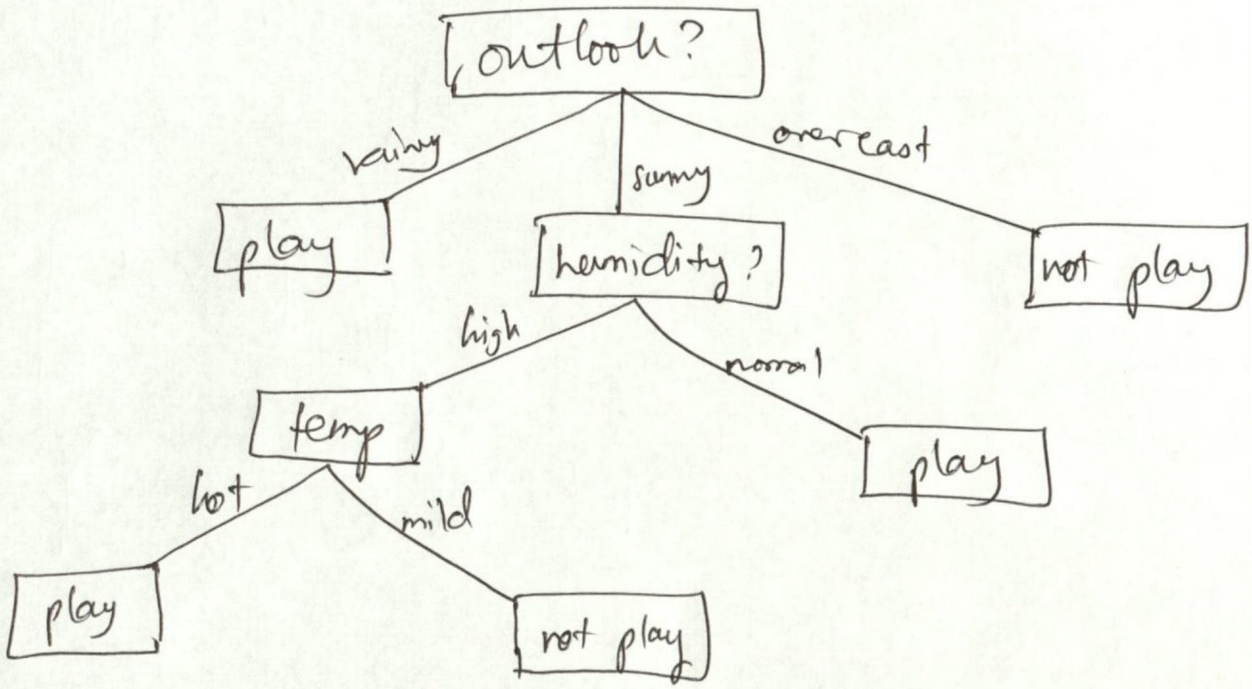
Step-2: –

The splits for humidity has the smallest gini index; so the tree becomes:



Step-3:-

The multi-way split of {hot, mild, cool} has the smallest weighted gini index (0.20) So, the final tree is:-



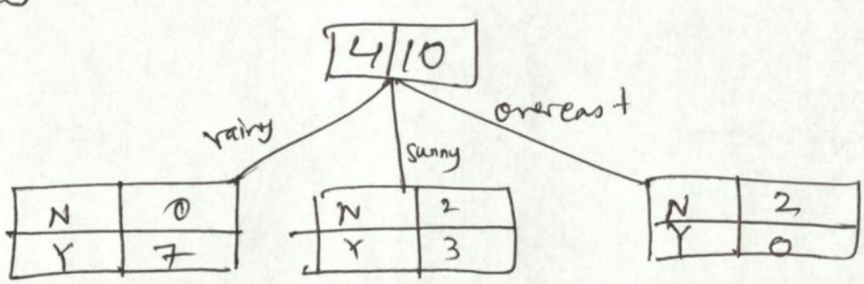
Ans. to 1(a)

Entropy

Step-1 :-

The entropy for the two classes initially is 1.052); a split that maximizes gain, will be selected

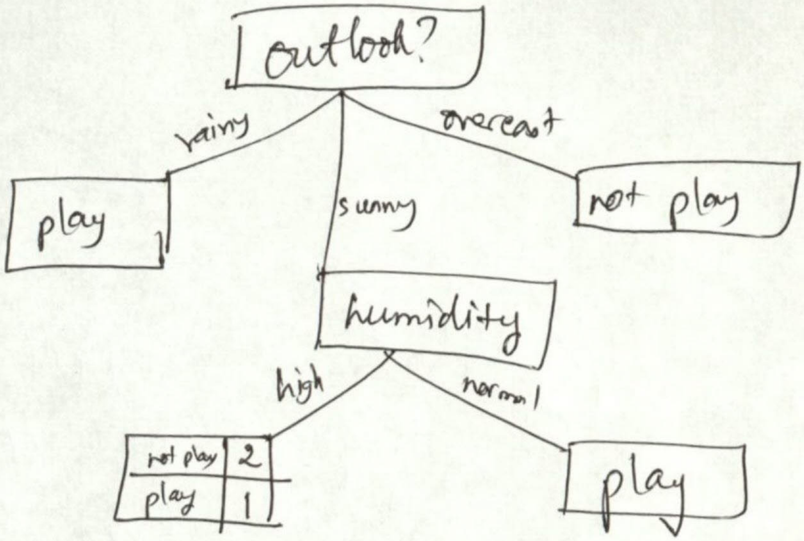
For attribute outlook, with splits {overcast, sunny, rainy} gives the lowest entropy than other splits (0.528792), such as {overcast + rainy, sunny}, {sunny + overcast, rainy} and {overcast, rainy +sunny}. So the tree becomes



step-2:-

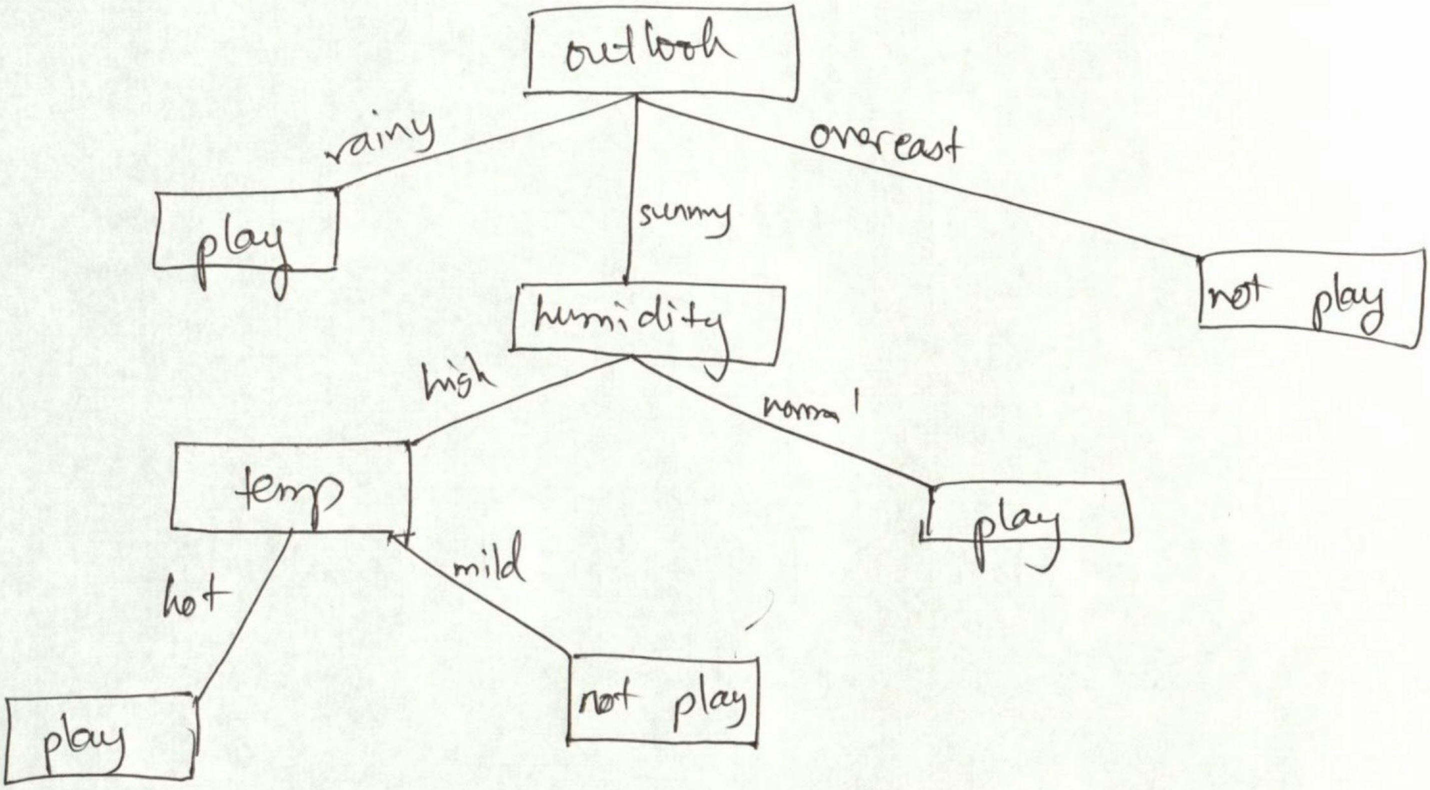
For attribute humidity we have ~~one~~ possible split that has the entropy of 0.91829

so, the tree becomes



Step-3:-

For attribute temp, the multi-split gives the smallest entropy and the final tree is

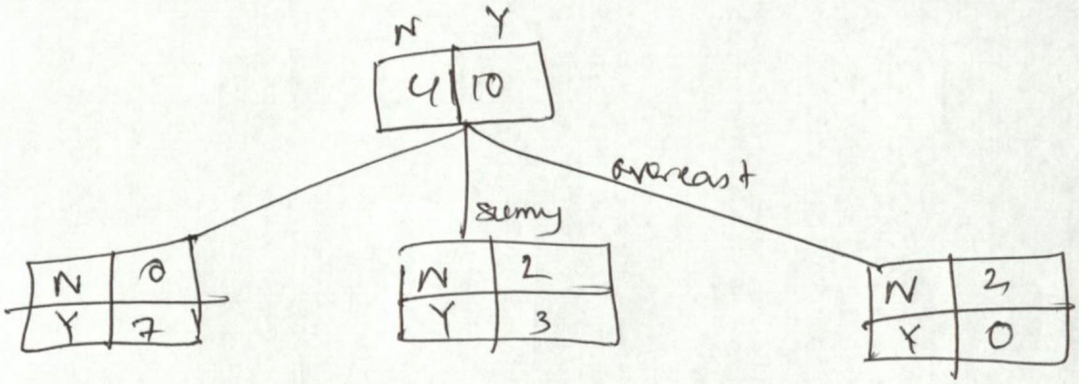


Ans. to Q 1(a)

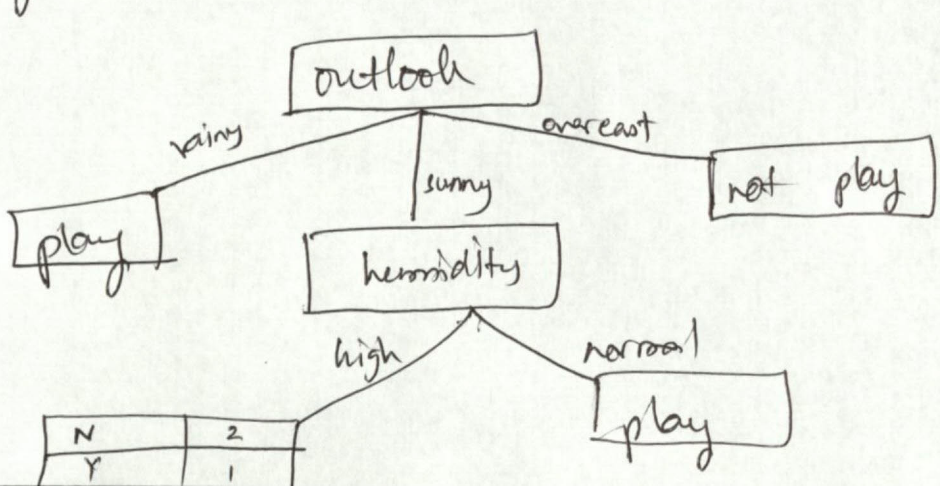
classification error:-

Step-1:-
For attribute outlook, with splits {overcast, sunny, rainy} gives the smallest classification error (0.07142) from other splits such as {overcast + sunny, rainy}, {overcast, sunny + rainy}, and {sunny, overcast + rainy}.

The tree becomes

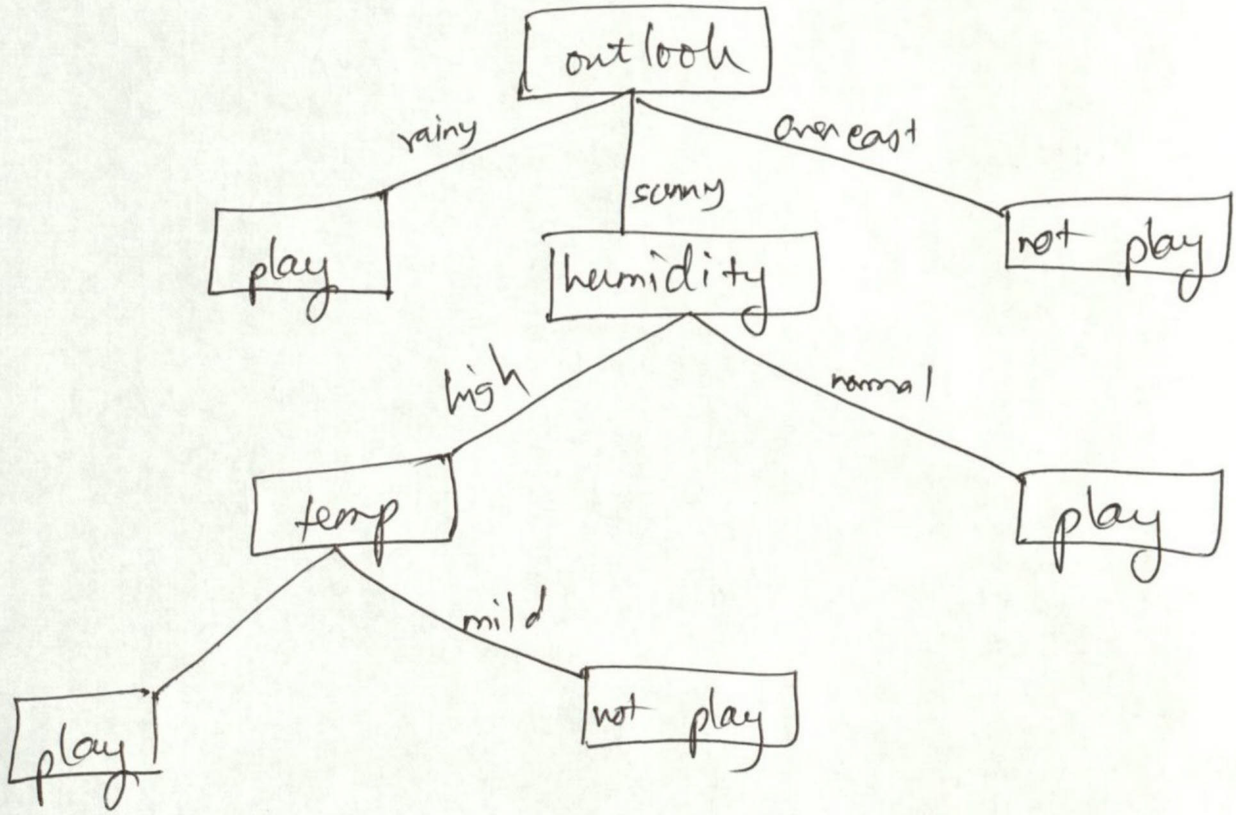


Step-2:- For attribute humidity we have one possible split that has the classification error of $1/5$, so the tree becomes,



Step-3:-

For attribute temp, the multi-split gives the smallest classification and the final tree is



Ans. to. 1(b)

indi. class accuracy for 'play' = $\frac{10}{10}$ = 100%

indi. .. .. 'not play' = $\frac{3}{4}$ = 75%

overall class accuracy = $\frac{13}{14}$ = 92.85%

Ans to. 1(c)

	Gini index	Entropy	Mis-classification error
1	Not play	Not play	Not play
2	Not play	Not play	Not play
3	Play	play	play
4	play	play	play
5	play	play	play

Ans. to. 2(a)

optimistic generalization error rate $= \frac{26}{100} = 0.26$

pessimistic " " " $= \frac{26 + 6 \cdot 12}{100} = 0.38$

Ans. to 2(b)

Error rate for test set $= \frac{21}{50} = 0.42$

Ans. to 2(c)

optimistic generalⁿ error rate $= \frac{32}{100} = 0.32$

pessimistic " " " $= \frac{32 + 4 \times 2}{100} = 0.40$

error rate for test set $= \frac{14}{50} = 0.28$

Ans to Q6:-

Three circles were misclassified as squares
One square was " " circle

Error rate = $\frac{3+1}{20} = 20\%$

Ans to Q3:-

a. $P(play = yes) = \frac{10}{14}$, $P(play = no) = \frac{4}{14}$

as, $P(play = yes) > P(play = no)$, will play

b.

$P(play = yes \mid hemidity = normal \wedge temp = mild \wedge outlook = sunny)$

$= P(hemidity = normal \mid play = yes) * P(temp = mild \mid play = yes)$
$* P(outlook = sunny \mid play = yes)$

$= \frac{6}{10} * \frac{4}{10} * \frac{3}{10} = 0.072$

$P(play = no \mid hemidity = normal \wedge temp = mild \wedge outlook = sunny)$

$= P(hemidity = normal \mid play = no) * P(temp = mild \mid play = no) *$
$P(outlook = sunny \mid play = no) = \frac{1}{4} * \frac{2}{4} * \frac{2}{4} = 0.0625$

as the probability score is higher for playing,
so will play

c. $$P(play=yes \mid temp=mild) = \frac{P(temp=mild \mid play=yes) * p(play=yes)}{P(temp=mild)}$$

$$= \frac{\frac{4}{10} * \frac{10}{14}}{6/14} = 0.47619$$

$$P(play=no \mid temp=mild) = \cancel{0.3}\ \frac{P(temp=mild \mid play=no) * P(play=no)}{P(temp=mild)}$$

$$= 0.333$$

as the prob. seen is higher for playing, will play.

d. $P(play=yes \mid temp=mild \wedge humidity=high)$
$= P(temp=mild \mid play=yes) * \cancel{p} P(humidity=high \mid play=yes)$
$= 0.16$
$P(play=no \mid temp=mild \wedge humidity=high)$
$= P(temp=mild \mid play=no) * P(humidity=high \mid play=no)$
$= \frac{2}{4} * \frac{3}{4} = 0.375$

as prob. for not playing is higher, will not play

e. $P(play=yes \mid outlook=overcast \wedge$
$temp=mild \wedge humidity=high)$

$= P(\text{overcast} \mid \text{play}=\text{yes}) * P(\text{mild} \mid \text{play}=\text{yes}) * P(\text{humidity}=\text{high} \mid \text{play}=\text{yes})$

$= \frac{0}{10} * \frac{4}{10} * \frac{4}{10} = 0$

$P(\text{play}=\text{no} \mid \text{outlook}=\text{overcast} \wedge \text{temp}=\text{mild} \wedge \text{humidity}=\text{high})$

$= P(\text{outlook}=\text{overcast} \mid \text{play}=\text{no}) * P(\text{mild} \mid \text{play}=\text{no}) * P(\text{humidity}=\text{high} \mid \text{play}=\text{no})$

$= \frac{2}{4} * \frac{2}{4} * \frac{3}{4} = 0.1875$

Will not play, as prob. for not playing is higher.

As there were no entries for outlook = overcast for play = yes, the conditional prob. was zero; one way to solve this is to use Laplace (add-one) smoothing to the numerator

$= P(\text{overcast} \mid \text{play}=\text{yes}) * P(\text{mild} \mid \text{play}=\text{yes}) * P(\text{humidity}=\text{high} \mid \text{play}=\text{yes})$

$= \frac{0}{10} * \frac{4}{10} * \frac{4}{10} = 0$

$P(\text{play}=\text{no} \mid \text{outlook}=\text{overcast} \wedge \text{temp}=\text{mild} \wedge \text{humidity}=\text{high})$

$= P(\text{outlook}=\text{overcast} \mid \text{play}=\text{no}) * P(\text{mild} \mid \text{play}=\text{no}) * P(\text{humidity}=\text{high} \mid \text{play}=\text{no})$

$= \frac{2}{4} * \frac{2}{4} * \frac{3}{4} = 0.1875$

Will not play, as prob. for not playing is higher.

As there were no entries for outlook=overcast for play=yes, the conditional prob. was zero; one way to solve this is to use Laplace (add-one) smoothing to the numerator

Ans- to Q4

attribute temp:-

	play = yes	play = no
hot	3	1
cool	3	1
mild	4	2

R1:-

if temp = hot then play
elseif temp = cool then play
elseif temp = mild then play

accuracy = 0.71428

attribute:- outlook:-

R2:-

	play = yes	play = no
sunny	3	2
overcast	0	2
rainy	7	0

if outlook = sunny, then play
elseif outlook = overcast, then dont play
elseif outlook = rainy, then play

accuracy = 0.85714

attribute = humidity

if humidity = high then play
" " = normal then play

accuracy = 0.71428

The best attribute is 'outlook', as it has the smallest total error rate = 0.14286

The best attribute is 'outlook', as it has the smallest total error rate = 0.14286

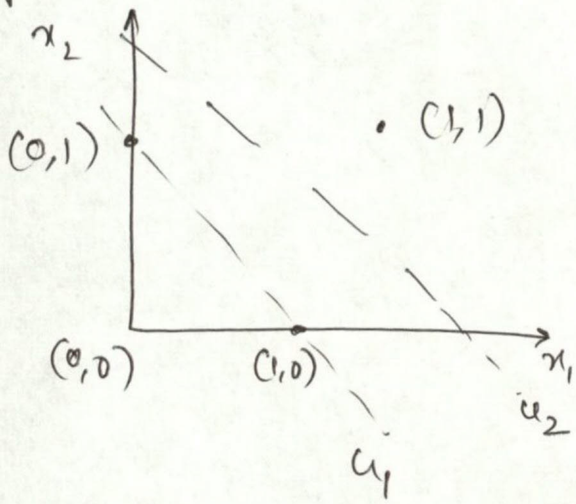
The best attribute is 'outlook', as it has the smallest total error rate = 0.14286

The best attribute is 'outlook', as it has the smallest total error rate = 0.14286

Ansto. Q5:-

One added neuron will solve the XOR problem

Hyperplanes:-



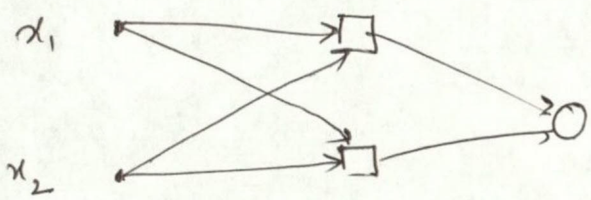
for 1^{st} neuron:

$$u_1 = b_1 + w_{11} x_1 + w_{12} x_2 > 0$$

for 2nd neuron:-

$$u_2 = b_2 + w_{21} x_1 + w_{22} x_2 < 0$$

network:-



weights:-

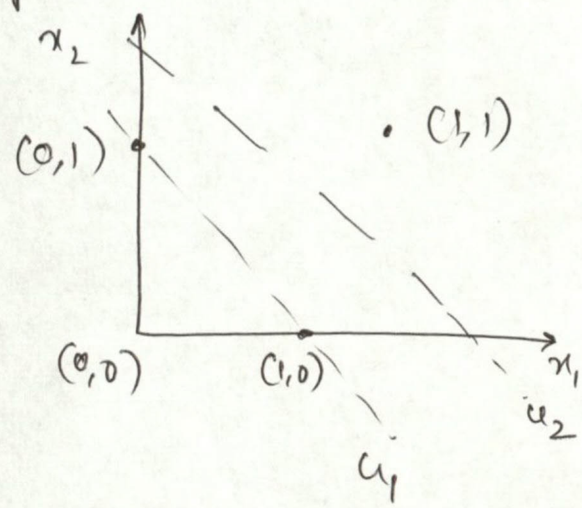
$w_{11} = -1$ $\quad w_{21} = 1$ $\quad b_1 = -1/2$

$w_{21} = 1$ $\quad w_{22} = -1$ $\quad b_2 = -1/2$

Ansto. Q5:-

One added neuron will solve the XOR problem

Hyperplanes:-



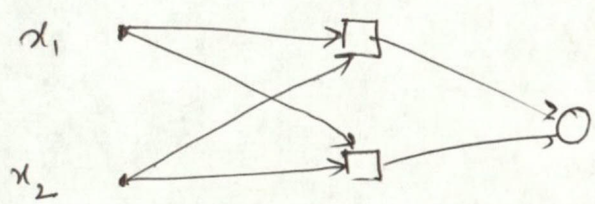
for 1st neuron:

$$u_1 = b_1 + w_{11} x_1 + w_{12} x_2 > 0$$

for 2nd neuron:-

$$u_2 = b_2 + w_{21} x_1 + w_{22} x_2 < 0$$

network:-



weights:-

$w_{11} = -1$ $\quad$ $w_{21} = 1$ $\quad$ $b_1 = -1/2$

$w_{21} = 1$ $\quad$ $w_{22} = -1$ $\quad$ $b_2 = -1/2$