CS 6375
ASSIGNMENT _____1_____

Names of students in your group:

Amandeep Singh(axs170032)

Number of free late days used: _____0_____

Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

Sol 2 False negative cases: 20%
False positive Cases: 10%

Sol 3. a) Specific hypothesis:

Pros: Specific hypothesis is always consistent with the training data.

Cons: It gives us false negative.

(b) General hypothesis:

Pros: Useful when one hypothesis is more general than others, so More study number of training examples can be used.

Cons: General hypothesis gives more false positive

Sol 4. Consistent hypothesis: If a hypothesis evaluates "T" for all positive example and "F" for all negative example, ie hypothesis is able to classify the example in the the training data, then it is said to be consistent.

Version Space: with respect to hypothesis space (H) and training examples (D), is the subset of hypothesis from H, consistent with the training example in D.

Sol 5 The most general hypothesis has "?" as don't care value for each attribute.

Sol 7.

<1, 1, 0, 1, 1>	1	D_1
<0, 1, 0, 1, 1>	0	D_2
<1, 1, 1, 1, 0>	1	D_3
<0, 0, 0, 1, 1>	0	D_4
<1, 1, 1, 1, 1>	1	D_5

In Find-S algorithm, we go from most specific hypothesis to general. and consider only (+)ve example.

$S \Rightarrow \langle \phi, \phi, \phi, \phi, \phi \rangle$

For D1

$S \Rightarrow \langle 1, 1, 0, 1, 1 \rangle$

For D3

$S \Rightarrow \langle 1, 1, ?, 1, ? \rangle$

For D5

$S \Rightarrow \langle 1, 1, ?, 1, ? \rangle$

Ans $S \Rightarrow \langle 1, 1, ?, 1, ? \rangle$

Sol (8) For this decision tree, we'll consider following attribute.

X_{GPA}

$X_{GPA} \geq 3.5$, $X_{EXP} \geq 3$, $X_{EXP} \geq 1$

DNF for this decision tree will be.

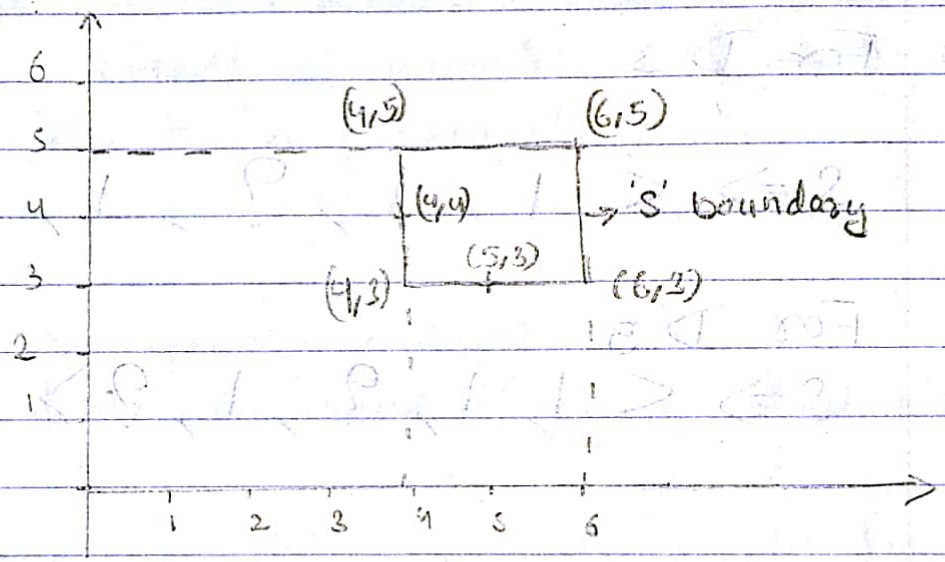
$$\{ (\neg X_{GPA} \geq 3.5 \wedge X_{EXP} \geq 3) \vee (X_{GPA} \geq 3.5 \wedge X_{EXP} \geq 1) \}$$

Sol 9.

From TOM MITCHELL

Q2.4

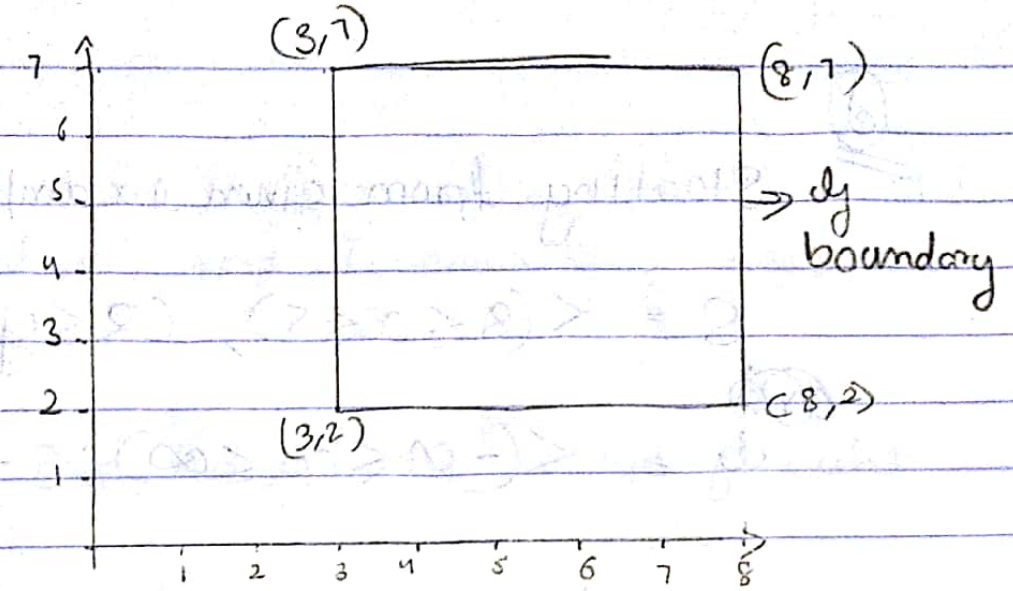
(a)



~~C.G.4~~

$$S = \{ < (4 \leq x \leq 6), (3 \leq y \leq 5) > \}$$

(b)



$S_1 \Rightarrow < (3 \leq x \leq 8), (2 \leq y \leq 7)$

(c) To reduce the size of version space, we need to increase the boundaries of 'S' and reduce the boundary of "S_1".
Therefore ~~our~~ query will be
$(x, y) \equiv (4, 6)$ either positive or negative.

Query that won't reduce version space will be within 'S' like (5,4) or beyond S_1 like (1,4).

(d) For algorithm to learn the target concept S boundary and H_2 boundary must coincide.

Therefore to achieve target concept

$S \Rightarrow (3 \leq x \leq 5)\ (2 \leq y \leq 9)$

from

$H_2 = (-\infty \leq x \leq \infty),\ (-\infty \leq y \leq \infty)$

we need minimum number of 4 example ie.

x	y	result
3	2	0 (-ve)
3	9	0
5	2	0
5	9	0

Q10

Initially we have.

$S = \langle (\phi, \phi, \phi, \phi), (\phi, \phi, \phi, \phi) \rangle$

$y = \langle (?, ?, ?, ?), (?, ?, ?, ?) \rangle$

(A) $D_1 \langle (UG, SE, l, hs), (gr, cs, h, hs) \rangle$ (+)

$S \Rightarrow \langle (ug, se, l, hs), (gr, cs, h, hs) \rangle$

$y = \langle (?, ?, ?, ?), (?, ?, ?, ?) \rangle$

(A) $D_2 \langle (ug, se, h, fr), (gr, cs, h, hs) \rangle$ (+)

$S \Rightarrow \langle (ug, se, ?, ?), (gr, cs, h, hs) \rangle$

$y = \langle (?, ?, ?, ?), (?, ?, ?, ?) \rangle$

(A) $D_3 \langle (gr, se, l, so), (gr, cs, h, se) \rangle$ (-)

$S \Rightarrow \langle (ug, se, ?, ?), (gr, cs, h, hs) \rangle$

g ⇒ {<(ug, ?, ?, ?), (?, ?, ?, ?)>,
<(?, ?, ?, ?), (?, ?, ?, hs)>}

D_4 = <(ug, se, d, jo), (gr, sa, h, ju)> +

~~S4~~ S ⇒ <(ug, se, ?, ?), (gr, ?, h, ?)>

dy ⇒ <(ug, ?, ?, ?), (?, ?, ?, ?)>

(b) <(ug, se, ?, ?), (gr, ?, h, ?)>
<(ug, se, ?, ?), (gr, ?, ?, ?)>
<(ug, se, ?, ?), (?, ?, h, ?)>
<(ug, se, ?, ?), (?, ?, ?, ?)>

<(ug, ?, ?, ?), (gr, ?, h, ?)>
<(ug, ?, ?, ?), (gr, ?, ?, ?)> — ①
<(ug, ?, ?, ?), (?, ?, h, ?)>
<(ug, ?, ?, ?), (?, ?, ?, ?)> — ②

Tatal hypothesis: 8

given

And only 2 satisfies data point

Sol(6) If all attribute are included with value

(a) 0 or 1 ∴ No. of possible values = 2

No. of instances possible for 1x1

$= 2 \times 2 \times 2 \times 2$

$= \boxed{16}$

(b) We can choose +ve literal, -ve literal or no literal at all i.e '?'

∴ No. of hypothesis $= 3^4$

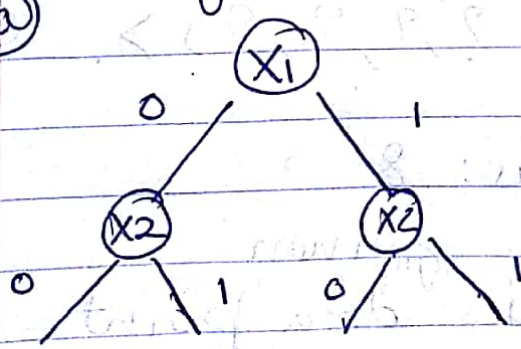
$= 81$

(c) No. of hypothesis $= 3^4$

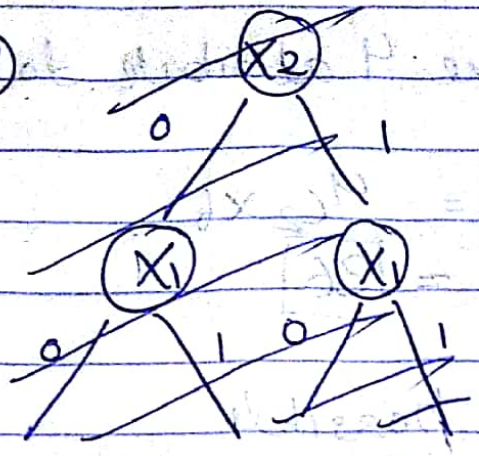
$= 81$

(d) From 2 attribute, we can make 6 decision tree of depth 2. as follows

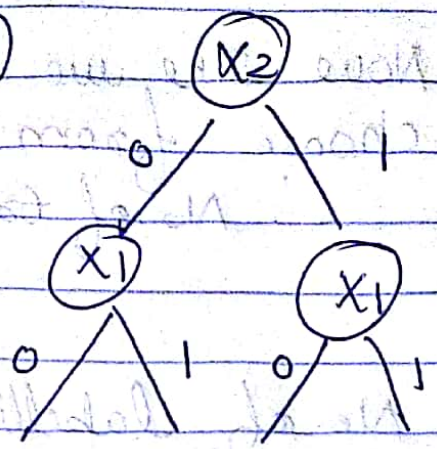
(a)



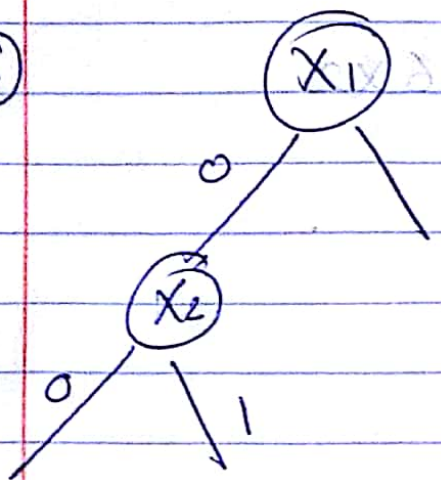
b)



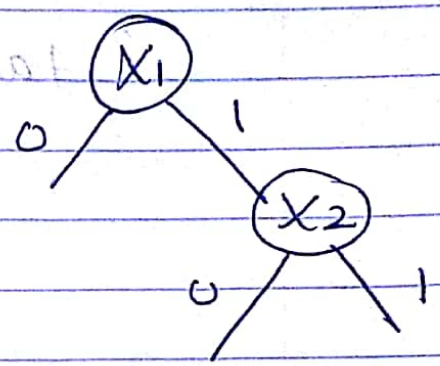
b)



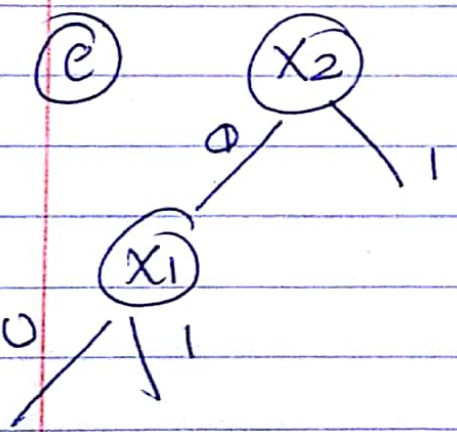
c)



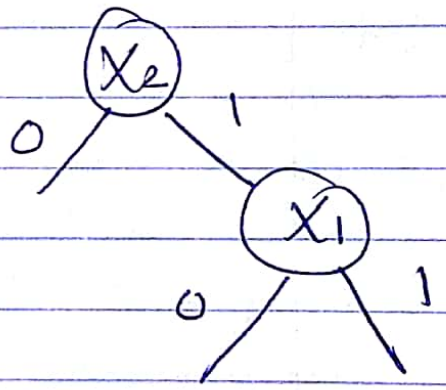
d)



e)



f)



Now since we have 4 attribute to choose from

$\therefore$ No. of trees $= {}^4C_2 \times 6$

$= \boxed{36}$

(e) No. of labelling possible (+ve, -ve) or (-ve, +ve)

$\therefore$ total $= {}^4C_2 \times 6 \times 2$

$= \boxed{72}$