Q a) Find S &

Time	Weather	Temp.	Company	Humidity	Wind	Goes
Morning	Sunny	Warm	Yes	Mild	Strong	Yes
Evening	Rainy	Cold	No	Mild	Normal	No
Morning	Sunny	Moderate	Yes	Normal	Normal	Yes
Evening	Sunny	Cold	Yes	High	Strong	Yes

Initialise most specific hypothesis
$S_0 = \langle \phi, \phi, \phi, \phi, \phi, \phi \rangle$

First example is +ve. So compare with it
$S_1 = \langle$ Morning, Sunny, Warm, Yes, Mild, Strong $\rangle$

Second example is -ve, so do nothing
$S_2 = \langle$ Morning, Sunny, Warm, Yes, Mild, Strong $\rangle$

Third is again +ve, so compare with it
$S_3 = \langle$ Morning, Sunny, ?, Yes, ?, ? $\rangle$

Fourth example is +ve, so again compare
$S_4 = \langle$?, Sunny, ?, Yes, ?, ? $\rangle$

Advantages & The Find-S algorithm only considers the +ve examples and eliminate -ve examples. For each +ve examples, the algorithm checks for each attribute in the example. If attribute value is same then algorithm moves but if not same then changes to '?'

Limitations & i) There is no way to determine if

The hypothesis is consistent throughout the data

:) Inconsistent training set can actually mislead the Find-S algorithm, since it ignores the -ve examples.

① b) Candidate Elimination &

Example

Citations	Size	InLibrary	Price	Editions	Buy
Some	Small	No	Affordable	One	No
Many	Big	No	Expensive	Many	Yes
Some	Big	Always	Expensive	Few	No
Many	Medium	No	Expensive	Many	Yes
Many	Small	No	Affordable	Many	Yes

Initialize most specific and most general hypothesis

S_0: ⟨ϕ, ϕ, ϕ, ϕ, ϕ, ϕ⟩

S_1: ⟨ϕ, ϕ, ϕ, ϕ, ϕ, ϕ⟩

S_2: ⟨Many, Big, No, Exp, Many⟩

S_3: ⟨Many, Big, No, Exp, Many⟩

G_0: ⟨?, ?, ?, ?, ?, ?⟩

G_1: ⟨Many, ?, ?, ?, ?, ?⟩ ⟨?, Big, ?, ?, ?, ?⟩
⟨?, Medium, ?, ?, ?, ?⟩ ⟨?, ?, Always, ?, ?⟩
⟨?, ?, ?, Exp, ?, ?⟩ ⟨?, ?, ?, ?, Many⟩
⟨?, ?, ?, ?, Few⟩

G_2: ⟨Many, ?, ?, ?, ?, ?⟩ ⟨?, Big, ?, ?, ?, ?⟩
⟨?, ?, ?, Exp, ?, ?⟩ ⟨?, ?, ?, ?, Many⟩

G_3: ⟨Many, ?, ?, ?, ?, ?⟩ ⟨Many, Big, ?, ?, ?, ?⟩
⟨?, Big, No, ?, ?, ?⟩ ⟨?, Big, ?, Always, ?, ?⟩
⟨?, Big, ?, ?, Many⟩ ⟨Many, ?, ?, Exp, ?, ?⟩
⟨?, Medium, Exp, ?, ?, ?⟩ ⟨?, ?, No, Exp, ?, ?⟩
⟨?, ?, ?, Exp, Many⟩ ⟨?, ?, ?, ?, Many⟩

S_4 : < Many, ?, NO, Exp, Many >	G_4 : < Many, ?, ?, ?, ? >, < Many, ?, Exp, ? > < ?, Medium, ?, Exp, ? >, < ?, NO, Exp, ? > < ?, ?, ?, Exp, Many >, < ?, ?, ?, ?, Many >
S_5 : < Many, ?, NO, ?, Many >	G_5 : < Many, ?, ?, ?, ? >, < ?, ?, ?, ?, Many >

∴ S : < Many, ?, NO, ?, Many >

G : < Many, ?, ?, ?, ? > < ?, ?, ?, ?, Many >

(2) a)

Supervised	Unsupervised
i) They are trained using labeled data.	i) They are trained using unlabeled data.
ii) It takes direct feedback to check if it is predicting correct output or not.	ii) It does not take any feedback.
iii) It predicts the output	iii) It finds the hidden patterns in data.
iv) In this input data is provided to the model along with the output.	iv) In this, only input data is provided to the model
v) The goal is to predict the output when it is given new data.	v) The goal is to find the hidden patterns and useful insights from unknown dataset.
vi) It produces an accurate result.	vi) It may give less accurate result as compared to supervised learning.
vii) It includes Linear & Logistic Regression, SVM, Decision Tree, Bayesian Logic etc.	vii) It includes various algorithm such as Clustering, KNN, etc

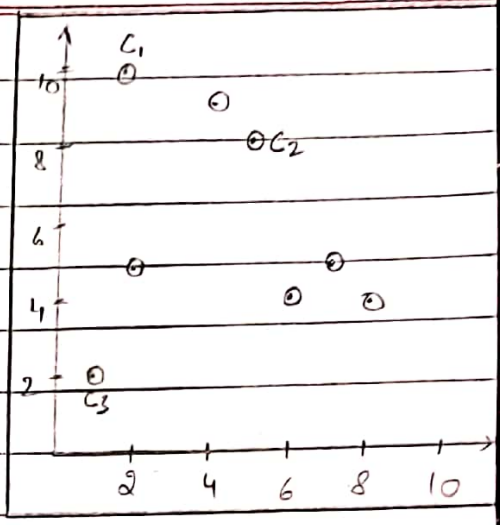
(2) b) We are solving this question by KNN in which distance formula is $\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$

Height	Weight	Distance	Size
158	58	4.24	M
158	59	3.60	M
158	63	3.60	M
160	59	2.23	M
160	60	1.41	M
163	60	2.23	M
163	61	2	M
160	64	3.16	L
163	64	3.60	L
165	61	4	L
165	62	4.12	L
165	65	5.66	L
168	62	7.07	L
168	63	7.28	L
168	66	8.60	L
170	69	9.21	L
170	64	9.49	L
170	68	11.40	L

Now, we have to find the size for height 161 & weight 61 $\therefore x_2 = 161$ & $y_2 = 61$
By putting the value of x_2 & y_2 in every row of above table calculate the distance

Suppose the value of k is 3 then least 3 distances are 1.41, 2, 2.23
$\therefore$ Size is M

(2) c) As the value of k is 3 and three centroides are $C_1(2,10)$, $C_2(5,8)$, $C_3(1,2)$



Distence = $|x_2 - x_1| + |y_2 - y_1|$

Now, we have to find the distence of every point from three centroids

x	2	2	8	5	7	6	1	4
y	10	5	4	8	5	4	2	9
$C_1(2,10)$	0	5	12	5	10	10	9	3
$C_2(5,8)$	5	6	7	0	5	5	10	2
$C_3(1,2)$	9	4	9	10	9	7	0	10
Cluster	C_1	C_3	C_2	C_2	C_2	C_2	C_3	C_2

Now calculate the new centroids

$C_1 = (2,10)$

$C_2 = (8,4)(5,8)(7,5)(6,4)(4,9)$

$x = \frac{8+5+7+6+4}{5} = 6$, $y = \frac{4+8+5+4+9}{5} = 6$

new $C_2 = (6,6)$

$C_3 = (2,5)(1,2)$

$x = \frac{2+1}{2} = 1.5$, $y = \frac{5+2}{2} = 3.5$

new $C_3 = (1.5, 3.5)$

x	2	2	8	5	7	6	1	4
y	10	5	4	8	5	4	2	9
$C_1(2,10)$	0	5	12	5	10	10	9	3
$C_2(6,6)$	8	5	4	3	2	2	9	5
$C_3(1.5,3.5)$	7	2	7	8	7	5	2	7
Cluster	C_1	C_3	C_2	C_2	C_2	C_2	C_3	C_1

Now, again calculate the new centroids.

$C_1 = (2,10), (4,9)$

$x = \frac{2+4}{2} = 3$, $y = \frac{10+9}{2} = 9.5$

new $C_1 = (3, 9.5)$

$C_2 = (8,4), (5,8), (7,5), (6,4)$

$x = \frac{8+5+7+6}{4} = 6.5$, $y = \frac{4+8+5+4}{4} = 5.25$

new $C_2 = (6.5, 5.25)$

$C_3 = (2,5), (1,2)$

$x = \frac{2+1}{2} = 1.5$, $y = \frac{5+2}{2} = 3.5$

new $C_3 = (1.5, 3.5)$

Now, the three cluster centres after the second iteration is

$C_1(3, 9.5)$, $C_2(6.5, 5.25)$, $C_3(1.5, 3.5)$

① a) Appropriate & i) Virtual Assistance
ii) Self Driven Vehicles
iii) Facial Recognition

Inappropriate & i) Execute a query to database
ii) Use word
iii) Calculate Payroll