CS 6375 ASSIGNMENT 1

Names of students in your group:

- 1. Rutuja Kaushike (rnk170000)
- 2. Akhila Kancharana (axk180025)

CS 6375.502

Number of free late days used: 0

Note: You are allowed a <u>total</u> of 4 free late days for the <u>entire semester</u>. 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.

1. Machine Learning (Book by Tom Michelle)

```
Python code for question 1 is as follows:
```

```
def question1(theta0, theta1, alpha, examples, m):
    length=len(examples)
    for i in range(m):
        err= (theta0 + sum(((theta1*examples[i][0])-examples[i][1])**2 for i in
    range(length)))/(2*length)
        print("Theta 0 is ", theta0, "Theta 1 is ", theta1, "and Error is ",err)
        der1=[theta0 + (theta1*examples[i][0]) for i in range(length)]
        newTheta0=theta0-(alpha*sum(der1))/length
        der2=[theta1*examples[i][0] for i in range(length)]
        newTheta1=theta1-alpha*sum(der2)/length

        theta0=newTheta0
        theta1=newTheta1

examples=[(3,2),(1,2),(0,1),(4,3)]
question1(0,1,0.0245,examples,5)
```

The output is as follows:

```
>>>
```

===== RESTART: /Users/rutujakaushike/Documents/assignment1question1.py

Theta 0 is 0 Theta 1 is 1 and Error is 0.5

Theta 0 is -0.049 Theta 1 is 0.951 and Error is 0.4281782499999999

Theta 0 is -0.0943985 Theta 1 is 0.904401 and Error is 0.37450398610325

Theta 0 is -0.13640138575 Theta 1 is 0.860085351 and Error is 0.33670020754682795

Theta 0 is -0.175203733998125 Theta 1 is 0.8179411688010001 and Error is 0.3127338950087588

>>>

Hence, error goes down as iterations increase.

9							
Q	@2.						
9			test positive	fest vegative			
9	1000	Has disease	0.8	, 4			
9							
2				0.9			
		Doesn't have disease	Name of the state		A CONTRACTOR OF THE PARTY OF TH		
		•	THE CONTRACT OF THE CONTRACT O	2000 750	And the second s		
*		Here as given in t	he example.	a testing	The state of the s		
		Here, as given in the example, a testing, machine can identify the disease in 80% of the cases. i.e. true positive here is 80% i.e. 0.8.					
		the cases i.e. true	positive here	is 801, i.e. 0.	.8		
*		Also here true negative is 90% i.e.a machine					
	and an analysis of the second	is able to convertly medict for those who					
	end at the section of a page representation of the St.	do not have disease 90% of the time.					
•		- true negative is 90% i.e. 0.9.					
		Now as the positive	is 08. false	tve false			
-		Now, as true positive is 0.8, talse the false negative should be 0.2 which is 1-the positive.					
	Para Antonio III III III III III III III III III I	AS					
		True Positive = 1 - false Negative 0.8 = 1 - false negative i. false Negative is 0.2. i.e. 201. (20 percent)					
		0.8 = 1 - false	negative				
	LIGHT TO A STREET AND A STREET	: false Negative i	6 0.2 i.e.	201. (20 opslout	り		
	MALES AND THE STREET OF THE ST	Similarly,					
	<u> </u>	False polition = 1 -	- true reactive	e			
		false poeixve = 1 = 1	0,9		THE RESERVE OF THE PERSON OF T		
			0-1				
	ening programmer or manager	Hence, false positive		17. (10 percent)			
7		The state of the s		The second secon			

av no cumulator de	
Q3.	PROS AND CONS OF SPECIFIC AND GENERAL HYPO.
```	
a.	selecting the most specific hypothesis on a
Annacio e e e e e e e e e e e e e e e e e e e	Kaining data
Canana	
	PROS; CONS:
	1. Result is not 1. Avoids negative
, , , , , , , , , , , , , , , , , , ,	affected due to examples
Committee of the second se	the order of 2. No room for
na ye rama e e e mana e e e m	promples which generalization.
	we proceed.  3. It noise data
Annual of the second of the se	2. works fine with comes, result
	examples with gets affected.
	only positive data.
	3. Represents only
	Specific hypothesis
And the state of t	out gall consistent
manadaphy mine of hy minebody mines for homely homes had been early him and hy or additioned	oves. Only specific boundaries
A Security of the Security Sec	
6.	selecting the most general hypothesis (G) = based on a locaining data
The control of the co	hand on a learning data
WAS A STATE OF THE PARTY OF THE	PONE
	1 Washe on hatte positive 1. Noisy data
	and negative data misleads the
	2 motallested Result hypothesis
	is not affected due to (General Hypothesis)

the order of examples.

3. Finds most spe

general hypothesis
based on training data

2. Incoencert data leads

Q4. consistent hypothesis:

A hypothesis h is consistent with a set of training example D if and only if f(x) = c(x) for each example f(x) = f(x) in D.

Consistent  $(h, D) = (\forall (x, c(x)) \in D) h(x) = c(x)$ 

Version Space: The version space, denoted VSHD with respect to hypothesis space H and training examples D, is the subset of hypothesis from H consistent the training examples in D.

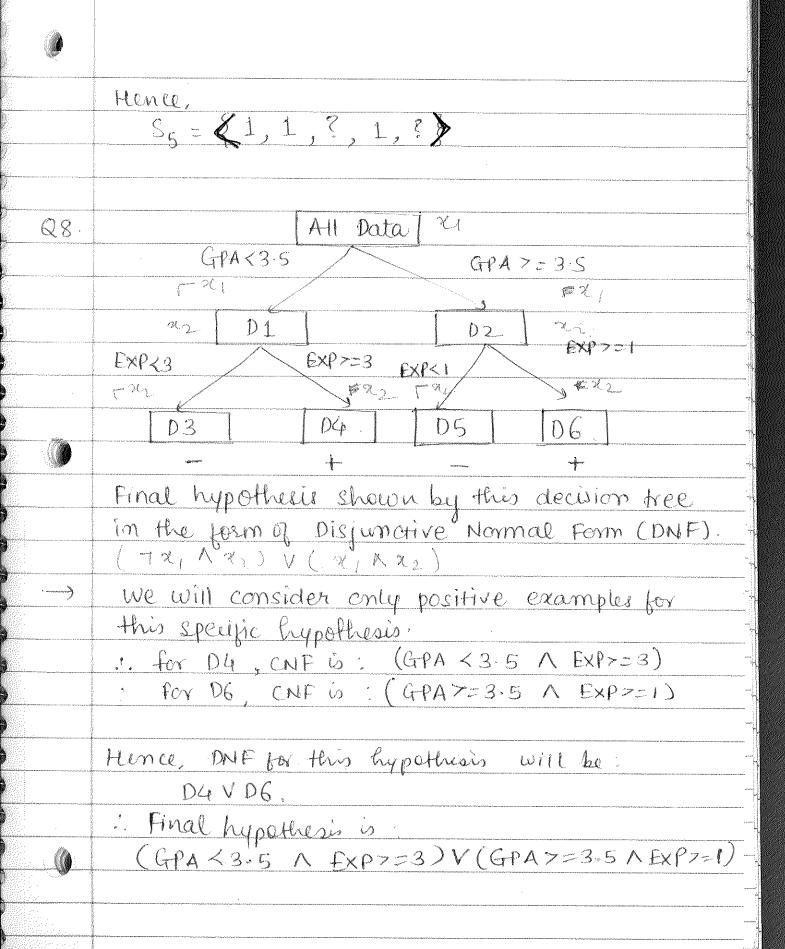
VSHID = {h & H | Consistent (h, D)}

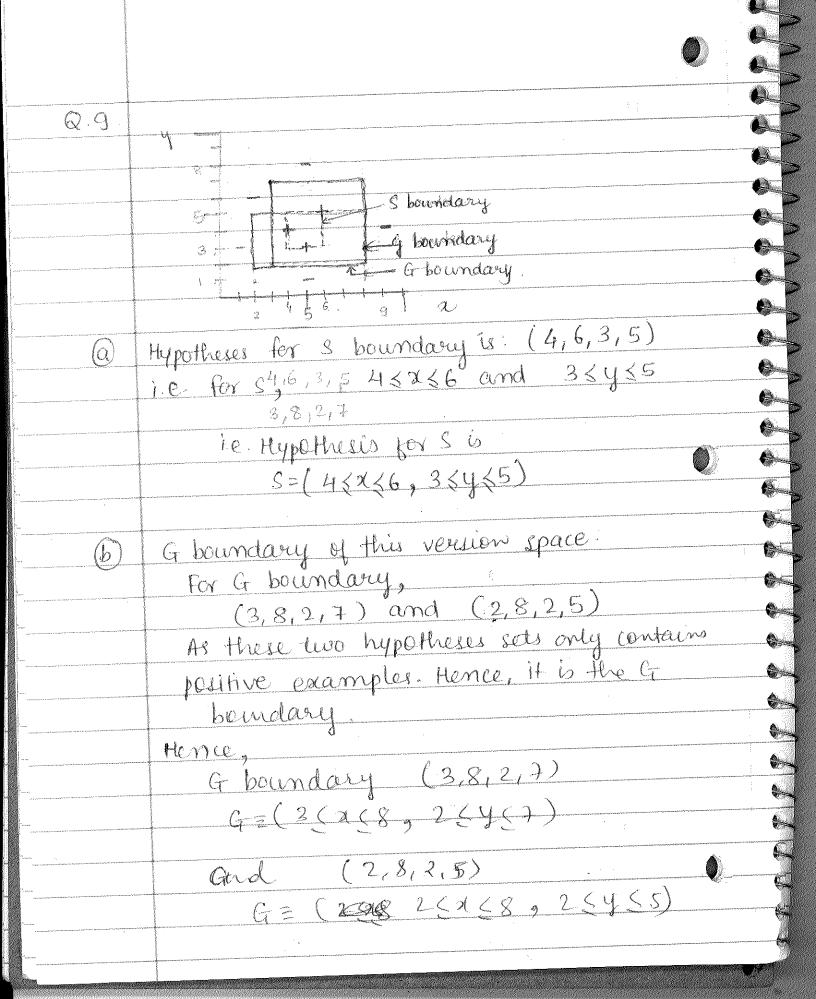
<i>:</i>	Abbaman il formiccione del				<b>4</b>
	icongine (di disensi) (di di				<b>Q</b> 2200
: : :	<b>A</b> D	The most general hypothesis? Value for each attribute.	hous		<u> </u>
ß ,	<b>Q</b> 5.	The most german requirements	Anna and a second		6
		? Value for each attribute.	MRYANIA MARKATI I AMARAMATAN MARKATI AMARKATAN MARKATI AMARKATAN MARKATI AMARKATAN MARKATAN MARKATAN MARKATAN M	transport of the state of the s	
4.,			Construction of the Constr		
				4.724.4	Parader P
			Antonialan yang ti i i i palamalyak yang yang pandi jana yang managan maliyak yang manay it	elektrika para samanda mpalika sana pari quantiya qila sama i	<b>-6</b> }
				amahalifiyyya madada ku iyya a masiliddiri	
	i de la compania del compania del compania de la compania del compa		Mayor	all appropries a manchane by the first paper on the artificial and all all artificials.	
	CALL + MINISTER CALL CALL CALL CALL CALL CALL CALL CAL				
	The second secon		y y met del fifth his last a start my my dy his trius a y m syn dy'n del de er start naphen sich die Victoria m		water to the state of the state
	The state of the s		معارفتها والمراس المتعدد بعادي والورث المتعارف المتعارف المتعارف المتعارف المتعارف المتعارف المتعارف والمتعارف		
	and control of the co				
	and the second s				
			nga manghana Baha Nga ya manghalin da Kata Sa Manarang na Madalanga ay mangang danda Nga Yanga danda Nga Manara	د المنظمة المنظ	6
			A STANDER COM TO THE STANDER OF THE	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	
			and the first of the first of the second of the first of	and the second s	6
	-		egypenne en fyr eildin in meg menni dyndysmigen og eildig fyll einnere en mylydyshal a fernion meg		
			and of the SEE of the sequence of the SEE of	one and the reason as all the second and all the se	ALLA SERVICIO DE LA COSTO DEL COSTO DE LA COSTO DEL COSTO DE LA COSTO DEL COSTO DE LA COSTO DE LA COSTO DE LA COSTO DEL COSTO DE LA COSTO DEL COSTO DE LA COSTO DE
	a contract of the contract of				
	And the second of the second o			AN AND A THE PARTY OF THE PARTY	<b>&amp;</b>
	A STATE OF THE PROPERTY OF THE			verson a a hijensiya ya wennaniya a diyahak kiki ka	
	t - di garantana Marayana da ana ana ana ana ana ana ana ana a			المستعمل والإنجاب والمستعمل والمستعمل والمستعمل والمستعمل والمستعمل والمستعمل والمستعمل والمستعمل والمستعمل وا	y 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
	To the second se		49 yekeyi zunammuş (Çiyayılan albaniya eş yekim a a kişişişiya keşamin kezeyinin kezeyinin kezeyinin kez		65-
	To 25 graphs to the second sec		Various administration as seminare and services as the animal objects as man developed by a seminary of c	z <b>O</b> lija	ß-
	and the second s		managhil let free comman eigeigight. In comman a to physic of the hand managet physical has a change follow th	m materialisti e e e e e e e e e e e e e e e e e e e	•
	to the state of th		ng-enging by partial a-scriptography ik ik ik is a quoq Senjilimma kronogisty-globilimik a urbiyiyiyi ikisi	ali inner e vergentiti kananan papa peretah inner e pereta	
	A 1 Clay of the State of the St		ne de autor y arrandi (AAN serroment des le les verques un mentre accessor a contradiga reminent de	the first section of a section of the section of th	
					W. 49 18 18 18 18 18 18 18 18 18 18 18 18 18

F: X74 X= (x1, x2, x3, x4) Q6 How many initances of i.e. IXI are possible? (a)x, x2, x3 and x4 are boolean values. Hence, assuming that each attended, number of instances = 2 for each attribute Hence = 2 x 2 x 2 x 2 = 24 = 16 instance. Let's assume that there are jour literals. (b): total instance space will be 24 = 16. Now, (for every literal, there is a positive i.e. attribute and a negation of that attribute: 2 for each. .. 2x2x2x2=24. for every attribute and negation of that, we have 2 values, 0 or 1 . .. we will have  $2^{121} = 2^{2} = 2^{16}$  in total. For each attribute, there are three labellings (c) i-e. (1,0,?) hence, the 1×1 i-e-instance space will be 3 for each attribute i.e. 34 in total. Now, for every hypothesis we propose, we have positive or a negative choice. i.e. again 2 choices for every hypothesis

Hence we will have  $2^{|X|} = 2^{34} = 81$   $2^{|X|} = 2^{|X|}$  in total. (d) Here, we have depth 2 and Lattributes. Hence, we can choose 40. 2 ow 94 atteibutes, for depth 2. Hence, for a combination problem (100 (order is not important), we will get  $4! = \mu \times 3 \times \ell \times 1 = 6. \text{ break}.$   $4 \cdot \ell = 2! \cdot 2! \quad 2 \times 2 \quad 2 \times 2$ If ordering is important and ifisdiff newly ordered tree is considered as new tree we will have  $4p = \frac{4!}{2!} = 124 \text{ ress. } \left[ np = \frac{h!}{(n-n)!} \right]$ (e) = 2x1x2x1 = 4 ways to choose. ba **6**>we have 6 distinct trees from *6*--Am (d), and we can label *6*-them in 4 diff ways 2 ways 2 way way Hence, total ways = 4 x L = 260. = 4 X6 = 26p.

Find 'S' algorithm. Q6 07. Let's Start "h" with  $S = \{ \Phi, \Phi, \Phi, \Phi, \Phi \}$ Example 1 is a positive example. After applying it en so, st becomes. (<1,1,0,1,1>,1) S1= \$1,1,0,1,1> example 2 is a negative example. : We Will exclude that : Spand So will be same.  $S_2 = \{1, 1, 0, 1, 1\}$ example 3 is a positive example, we will consider that : (21,1,1,1,0>,1)  $S_3 = \{1, 1, 2, 1, 2\}$ example 4 is a negative example. : we will not consider that: : 83 and 84 will be the Same : S4 = & 1, 1, 8, 1, 8 x example 5 is a positive example and we will consider it. (<1,1,1,1,17,1)





09 (c)Query to reduce the size of the vereion space, of If it is a -ve example, if we put it anywhere in between s and G boundary, it will reduce the hypothesis space of G' for example, if (5,6) is a -ve example, it will reduce the 'g' boundary d 10 this will reduce G' boundary. this will not charge S&G boundaries this will not change S&G boundaries 2 10 Now, for no change boundances, (i) if we define any positive point in between 'S' region / boundary, it will not change the existing 1512 G' boundaries for example, if (5,4) is a positive example, it won't change s& & boundaries. (ii) if we define any -ve point outside G boundary for example, (7,1) which is a - ve entry, & boundary and s boundary will not change.

(d) 3,9)0 (5,9) ( CECO ) 5 Let's take the target concept as (3 5255, 25y £9) the version space is as strown above. I for this, any two tve boundary points will suffice for perfect learning say, (3,2) and (5,9) are enough to maké positive boundances work. For negative points (G should exclude those), any 4 re points outride version Space will be enough i.e. (2,7), (4,1), (6,6) and (4,10) will be sufficient for drawing General Boundaries • Hence, any 2 positive points like (3,2), (5,9) • and 4 negative points like (2,7)(4,1)(6,6)(4,10) # i.e. 6 points min are required.

Q10.  $S_0 = \langle \{ \phi, \phi, \phi, \phi \} \langle \phi, \phi, \phi, \phi, \phi \rangle \rangle$ (a)  $G_0 = \langle (?,???) (?,?,?,?) \rangle$ first example is a positive one. :. S1 = ((ug, se, l, hs), (gr, cs, h, hs))  $G_1 = \langle (2, 2, 2, 2), (2, 2, 2, 2) \rangle$ second example is also a positive ove.  $S_2 = \langle (ug, se, ?, ?)(gr, cs, h, h, h) \rangle$  $G_2 = \langle (2, 2, 2, 2) (2, 2, 2, 2) \rangle$ third example is -ve though. : we won't change S2 for S3 :. S3 = < ( ug, se, ?, ?), (gr, cs, h, hs) > &  $G_3 = \langle (ug,?,?,?)(?,?,?,?)\rangle$  $\langle (1,?,?,?)(?,?,?,hs)\rangle$ Now, example 4 is +ve-: S4 = < (ug, se,?,?)(gr,?,h,?)>

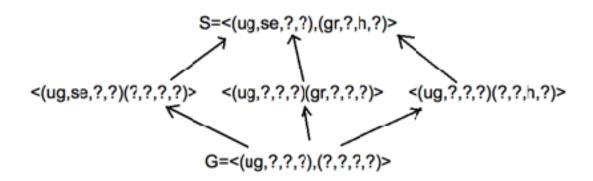
and Gy = < (Ug,?,?,?)(?,?,?,?)

10. b)

From 10.a) we have,

S = <(ug, se,?,?),(gr,?,h,?)>

G=<(ug,?,?,?),(?,?,?,?)>



Hence, there are total 6 consistent hypotheses present. Given hypotheses <(ug, cs, h, do), (gr, ma, l, se)> with class label '+' is then, consistent with three and inconsistent with three.