Assignment 1

Machine Learning CSC 354

BCS 7A/7B/7C

CLO1 [25 Marks]

Question 1   **[10 Marks]**

What is hypothesis space in machine learning. Consider dataset

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Cool	Change	Yes
2	Cloudy	Warm	Normal	Strong	Cool	Change	Yes
3	Rainy	Warm	Normal	Strong	Cool	Change	No

Explain why the size of the hypothesis space in the EnjoySport learning task is 973. How would the number of possible instances and possible hypotheses increase with the addition of the attribute WaterCurrent, which can take on the values Light, Moderate, or Strong? More generally, how does the number of possible instances and hypotheses grow with the addition of a new attribute A that takes on k possible values?

Question 2   **[15 Marks]**

Consider the instance space consisting of integer points in the x, y plane and the set of hypotheses H consisting of rectangles. More precisely, hypotheses are of the form $a \leq x \leq b, c \leq y \leq d$, where a, b, c, and d can be any integers.
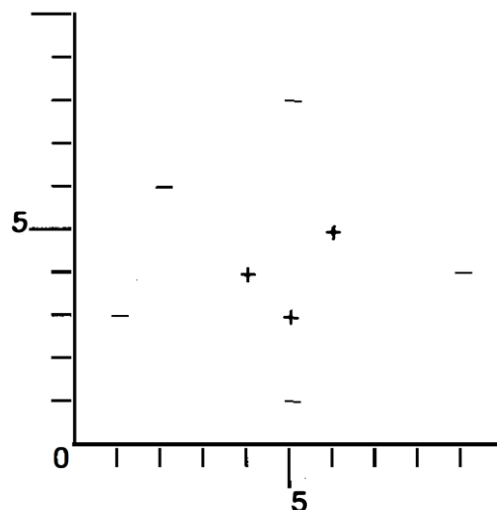
a. Consider the version space with respect to the set of positive (+) and negative, (-) training examples shown in Figure above. What is the S boundary of the version space in this case? Write out the hypotheses and draw them in on the diagram.
b. What is the G boundary of this version space? Write out the hypotheses and draw them in diagram.

Question 3   **[5 Marks]**

Give decision trees to represent the following Boolean functions:

a. $A \wedge \sim B$
b. $A \vee (B \wedge C)$
c. $A \otimes B$
d. $[A \wedge B] \vee [C \wedge D]$

CLO2 [20 Marks]

Question 4   **[10 Marks]**

Consider the following data and initial value of parameters compute the error and update of parameters using gradient descent algorithms for multiple linear regression model. $\theta_0 = 0.5, \theta_1 = 1.3, \theta_2 = -0.4, \theta_3 = -3$

X1	X2	X3	Y
2	3	8	6
3	5	7	3
1	4	9	7
6	7	3	1
7	2	2	4
4	1	1	5
5	5	3	6

Question 5   **[10 Marks]**

How kNN algorithm works? Provide the solution using kNN with $k = 11$, $k = 9$, $k = 7$ and k=5 for following data.

Test data

7	12	3	9	4	21	11

Training Data

x1	x2	x3	x4	x5	x6	x7	Class
2	2	13	12	5	6	6	0
3	21	7	1	13	7	1	1
18	15	3	2	19	6	21	1
16	16	15	12	5	17	4	1
11	13	1	1	14	20	14	1
3	19	1	7	20	1	10	1
3	4	3	3	12	19	20	0
2	6	19	2	17	17	10	0
20	5	12	2	13	15	15	0
6	15	16	2	6	12	8	1
20	15	1	16	17	5	4	1
14	16	3	20	8	15	10	0