Indian Institute of Technology Kharagpur
Department of Computer Science and Engineering

Machine Learning (CS60050)	**Mid-Semester Examination**	**Spring Semester, 2022-2023**
Date: 15-Feb-2023 (Wed, AN)	**Answer *all* questions.**	**Maximum Marks:** 60

Q1. [Concept Learning] **10 marks**

Consider the following set of attributes with their listed domain values:

- **Fever:** { *High, Moderate, None* }
- **Cough:** { *Intense, Mild* }
- **RunningNose:** { *Yes, No* }
- **Weakness:** { *Extreme, Slight* }
- **Headache:** { *Yes, No* }
- **Saturation:** { *Good, Bad* }

Suppose, based on the values of the above mentioned attributes, you are trying to learn the concept whether someone has **Covid** or not. You are given with the following training data set (4 examples):

Example	Fever	Cough	RunningNose	Weakness	Headache	Saturation	Covid
1	High	Mild	No	Extreme	No	Bad	*Yes*
2	High	Mild	No	Slight	No	Bad	*Yes*
3	None	Intense	No	Slight	No	Good	*No*
4	High	Mild	No	Extreme	Yes	Good	*Yes*

Consider the space H of conjunctive hypotheses, which, for each attribute, either:

- indicates by a '?' that any value is acceptable; or
- specifies a single required value (e.g., *Mild* for **Cough**); or
- indicates by a 'ϕ' that no value is acceptable.

Let a version space (a subset of consistent hypotheses in H) be represented by an S set (specific boundary, at the top) and a G set (general boundary, at the bottom). Suppose the 4 training examples above are presented in order. Answer the following.

(a) What is the total size (cardinality) of the possible hypothesis space? **(2)**

(b) Applying *Candidate-Elimination* algorithm, draw a diagram showing the evolution of the version space for concept **Covid** given the training examples, by clearly expressing S_1, G_1, S_2, G_2, S_3, G_3, S_4, G_4. If the G set does not change given a new example, just write $G_{i+1} = G_i$ $(1 \leq i < 4)$ next to the drawing of G_i (similarly for S set as well). **(4)**

(c) Write down all the hypotheses in the final version space (the ones that lie between S_4 and G_4 according to the partial ordering relation *Less-Specific-Than)*. **(2)**

(d) In the final version space, draw lines between hypotheses that are related by this relation. For example, there should be a line between $\langle ?, Mild, ?, ?, ? \rangle$ and $\langle ?, Mild, ?, Extreme, ?, ? \rangle$. **(2)**

Q2. [Decision-Tree Learning] **10 marks**

For a binary classification problem, consider the training examples shown in the following table.

Instance	A_1	A_2	A_3	Target Class
1	*True*	*True*	1.0	+
2	*True*	*True*	6.0	+
3	*True*	*False*	5.0	−
4	*False*	*False*	4.0	+
5	*False*	*True*	7.0	−
6	*False*	*True*	3.0	−
7	*False*	*False*	8.0	−
8	*True*	*False*	7.0	+
9	*False*	*True*	5.0	−

The attributes, A_1 and A_2, can take either *True* or *False* values, whereas A_3 is a *continuous* attribute. The **Target Class** can be either + (positive) or − (negative). Answer the following.

- **(a)** What is the entropy of this collection of training examples with respect to positive (+) class? **(2)**
- **(b)** What are the information gains of A_1 and A_2 relative to these training examples? **(3)**
- **(c)** For A_3, which is a continuous attribute, compute the information gain for every possible split. **(4)**
- **(d)** According to the information gain, which is the best split point considering only A_3 attribute? **(1)**
- **(e)** What is the best attribute (among A_1, A_2, A_3) to split according to the information gain? **(1)**

Q3. [Bayesian Leaning] **10 marks**

Consider the data set shown in the following table.

Instance	A	B	C	Class
1	0	0	1	−
2	1	0	1	+
3	0	1	0	−
4	1	0	0	−
5	1	0	1	+
6	0	0	1	+
7	1	1	0	−
8	0	0	0	−
9	0	1	0	+
10	1	1	1	+

The attributes, A, B and C, can take two values (either 1 or 0) and the **Class** can be either + or −. Answer the following.

- **(a)** Estimate the conditional probabilities for the following:
  $\mathbb{P}(A=0 \mid +),\ \mathbb{P}(B=1 \mid +),\ \mathbb{P}(C=1 \mid +),\ \mathbb{P}(A=0 \mid -),\ \mathbb{P}(B=1 \mid -),\ \mathbb{P}(C=1 \mid -)$. **(3)**
- **(b)** Use the conditional probabilities in part (a) to predict the class label for a given test sample, $(\mathsf{A}=0,\ \mathsf{B}=1,\ \mathsf{C}=1)$, using the Naive Bayes approach. **(4)**
- **(c)** Are the variables, A and B, independent with values, A = 1 and B = 1? **(1.5)**
- **(d)** Are these variables, A and B, conditionally independent with values, A = 1 and B = 1, given the class '+'? **(1.5)**

Q4. [Instance-based Learning] **6 marks**

Consider the one-dimensional data set shown in the following table.

x	0.5	3.0	4.5	4.6	4.9	5.2	5.3	5.5	7.0	9.5
y	−	−	+	+	+	−	−	+	−	−

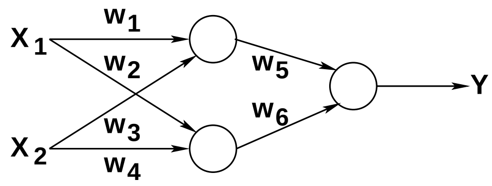
Here, *x* can take continuous values and *y* has two labels (+ and −). Answer the following.

- **(a)** Classify the data point $x = 5.0$ according to its 1-, 3-, 5-, and 9- nearest neighbors (using majority voting). Briefly explain your results. **(2)**
- **(b)** Again classify the same data point $x = 5.0$ according to its 1-, 3-, 5-, and 9- nearest neighbors (using distance-weighted voting). Briefly explain your results.
  Note: In distance-weighted scheme, the weights are inversely proportional to the Euclidean distances between two data points. **(4)**

Q5. [Perceptrons] **4 marks**

Suppose we have a multi-layer perceptron network (shown below) with linear activation units. In other words, the output of each unit is a constant *C* multiplied by the weighted sum of inputs.

Answer the following.



(a) Can any function that is represented by the above network also be represented by a single unit perceptron? If yes, draw the equivalent perceptron detailing the weights and the activation function. Otherwise, briefly explain why not possible. **(2)**

(b) Can the space of functions that is represented by the above network also be represented by linear regression? If yes, present the linear regression function detailing the coefficients. Otherwise, briefly explain why not possible. **(2)**

Q6. [Logistic Regression and Neural Network] **10 marks**

For a binary logistic regression model with input attribute set x and an output y, having an internal sigmoid activation function (of the form $\sigma(z) = \frac{1}{1+e^{-z}}$, where $z = w^T.x$ with weight vector w), we predict the output $y = 1$ when $\mathbb{P}(y = 1 \mid x\ ;\ w) \geq \frac{1}{2}$.

(a) Prove that, this logistic regression model is also a linear classifier. **(4)**

(b) Consider the XOR function $y = (x_1 \wedge \neg x_2) \vee (\neg x_1 \wedge x_2)$. We can alternatively express this as, $y = \begin{cases} \geq \frac{1}{2}, & \text{if } x_1 \neq x_2 \\ < \frac{1}{2}, & \text{otherwise} \end{cases}$. Using the above-mentioned binary logistic regrassion model as a unit having binary inputs $x_0 (= 1)$, x_1 and x_2 (i.e. $x = [1, x_1, x_2]^T$), and output y, draw a fully connected three-unit Neural Network that realizes the function $y = (x_1 \text{ XOR } x_2)$. Show the suitable weight vector, $w = [w_0, w_1, w_2]^T$, for each unit clearly. **(6)**

Q7. [Linear Classifier and Support Vector Machine] **10 marks**

Consider a set of 2-dimensional training data points (x_1, x_2) belonging to two classes '$+1$' and '-1', respectively, as shown below.

- Class '$+1$': $(3,1)$; $(3,-1)$; $(6,1)$; $(6,-1)$
- Class '-1': $(1,0)$; $(0,1)$; $(0,-1)$; $(-1,0)$

We design a linear hard-margin SVM to classify these linearly separable points. Answer the following.

(a) Pictorially (graphically) represent the constellation of data points and the optimal separating hyperplane. Write the equation of the optimal separator and mention the width of the margin (figuring it out manually from the diagram/graph you have shown). **(2)**

(b) Which data points are the support vectors here? **(2)**

(c) What weight vector and threshold (bias) value are being learnt using hard-margin SVM training algorithm with these eight training points? Show the detailed calculations. **(4)**

(d) Using the learnt weights and threshold values (in part (c)), what is the margin you get for the optimal classifier? Derive mathematically. **(2)**

— END —
