North East University Bangladesh
Department of Computer Science and Engineering
Program: BSc(Engg) in CSE
Semester Final Examination, Summer - 2022

Course Code: CSE 411 — Course Title: Artificial Intelligence
Time: 2 hours — Maximum Marks: 40

> Square bracketed numbers in the margin indicate marks for each part of a question.
> Boxed numbers in the margin indicate total marks of the question.

(Answer all of the following questions)

Question 1 .... Total Marks on Question 1 is: $\boxed{8}$

Answer **any 4 questions** from the following:

(a) Consider the following propositions:

$$a = \textit{The temperature reaches } 100 \textit{ degree.}$$
$$b = \textit{Water boils.}$$

Now convert the following sentences to compound propositions using logical connectives allowed in the syntax of propositional logic:

i. Water boils unless the temperature reaches 100 degree. [1]

ii. The temperature reaching 100 degree is necessary for boiling water. [1]

(b) Consider the same two propositions (a and b) as given in question **1.a**. Based on these, convert the following compound propositions to english sentences:

i. $a \leftrightarrow b$ [1]

ii. $a \leftarrow b$ [1]

(c) When is a logical proof called a sound proof? [2]

(d) What is the goal of supervised learning? [2]

(e) What is the reason for using regularizers in a supervised learning setting? [2]

Question 2 .... Total Marks on Question 2 is: $\boxed{12}$

Answer **any 4 questions** from the following:

(a) Let a and b be two atomic propositions. Now write the propositional semantics for the following three compound propositions:

i. $a \wedge b$ [1]

ii. $a \leftarrow b$ [1]

iii. $a \leftrightarrow b$ [1]

(b) For each of the following compound propositions, mention whether these are propositional definite clauses or not:

i. $p \leftarrow \neg q$ [1]

ii. q [1]

iii. $p \wedge q \leftarrow r$ [1]

(c) Write a pseudocode for the top-down proof on a knowledge base containing propositional definite clauses. [3]

(d) i. What is inductive bias? [1½]

ii. What is a bias feature? [1½]

(e) Consider the hypothesis space of a linear regression problem is restricted to have only integer parameter values. Would such a hypothesis space be easier to optimize or not? Justify your answer. [3]

Question 3 .... Total Marks on Question 3 is: $\boxed{20}$

Answer **any 4 questions** from the following:

(a) Consider you are given the following knowledge base: [5]

$p \leftarrow q.$
$\neg q.$
$\neg r \wedge p.$
$s \leftarrow t.$
$t.$

p	q	r	s	t
T	F	F	T	T
F	F	f	T	T
T	F	F		
F	F	F		

Write all possible models of this knowledge base. No explanation is required.

(b) Consider you are given the following knowledge base: [5]

$a \leftarrow b \wedge c.$
$a \leftarrow e \wedge f.$
$b \leftarrow f \wedge k.$
$c \leftarrow e.$
$d \leftarrow k.$
$e.$
$f \leftarrow j \wedge e.$
$f \leftarrow c.$
$j \leftarrow c.$

ai ← B

G ← [illegible]

Now find all the propositions entailed by this knowledge base by constructing a bottom-up proof. Show all steps of the proof along with your final result.

(c) Consider you are given the following dataset: [5]

x_1	x_2	t
2	4	6
6	36	42
4	16	20
3	9	12
5	25	30

1-min
6-2=9

Rewrite the same dataset by standardizing it. [Hint: $\mu = \frac{\sum_{i=1}^{N} x_i}{N}$ and $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)}{N}}$]

(d) Consider the same dataset given in question **3.c**. Rewrite the dataset by normalizing it in the range [0, 1]. [5]

(e) Consider you are given an N samples dataset with feature matrix $\boldsymbol{X} \in \mathbb{R}^{N \times D}$, target vector $\boldsymbol{t} \in \mathbb{R}^N$, and your goal is to learn a weight vector $\boldsymbol{w} \in \mathbb{R}^D$. A loss function for training over the dataset is defined as: [5]

$$\mathcal{L}(\boldsymbol{X}, \boldsymbol{t}, \boldsymbol{w}) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - y(\boldsymbol{x}_n, \boldsymbol{w})\}^2$$
$$= \frac{1}{2}\{\boldsymbol{t} - y(\boldsymbol{X}, \boldsymbol{w})\}^T\{\boldsymbol{t} - y(\boldsymbol{X}, \boldsymbol{w})\}$$

$X^T (X^T X)^{-1}$

where $y(\boldsymbol{X}, \boldsymbol{w}) = \boldsymbol{X}\boldsymbol{w}$ is defined as the hypothesis function of the regression problem. Note that, the first line of the loss function is in the summation form, and the second line is the same equation in the vector form.
Now your task is to **optimize this loss function** with respect to the parameter vector $\boldsymbol{w}$. If your solution is in closed form, then write the equation as it is. In other cases, write the equation as a batch gradient descent optimization equation.