CBCS SCHEME

USN | | | | | | | | | | |

18EC745

Seventh Semester B.E. Degree Examination, July/August 2022
Machine Learning with Python

Time: 3 hrs. Max. Marks: 100

Note: ***Answer any FIVE full questions, choosing ONE full question from each module.***

Module-1

1 a. Pick a learning task of playing chess. Describe it as precisely as possible with relevant machine learning notations. Represent the target function 'f' a linear combination of board features of your choice. Play it repeatedly against itself that uses a fixed evaluation function. Discuss the main tradeoffs you considered in designing this learning task. Plot the design choices you made. **(10 Marks)**

b. What are the elements of the version space? How are they ordered? What can be said about the meaning and sizes of S and G? Write G – to S ordering of hypotheses. **(10 Marks)**

OR

2 a. Implement Find-S and candidate elimination algorithms to solve the hypothesis of the given table and compare the outcome of hypotheses.

Origin	Manufacturing	Color	Decade	Type	Example
Japan	Honda	Blue	1980	Economy	+ve
Japan	Toyota	Green	1970	Sports	–ve
Japan	Toyota	Blue	1990	Economy	+ve
USA	Chrysler	Red	1980	Economy	–ve
Japan	Honda	White	1980	Economy	+ve
Japan	Toyota	Green	1980	Economy	+ve
Japan	Honda	Red	1990	Economy	–ve

(10 Marks)

b. Describe the propose and use of Numpy and matplotlib libraries. Mention any 5 function/modules used in both. Draw a line in a diagram from position (0, 0) to (6, 250) using Numpy and Matplotlib libraries. **(10 Marks)**

Module-2

3 a. Give decision tree representation of the following Boolean function :
i) $(A \wedge B) \vee (C \wedge D)$
ii) X xor Y. **(10 Marks)**

b. Mention any 5 problems/situations where decision tree is best suited. **(05 Marks)**

c. Write a short note on Occam's razor. **(05 Marks)**

OR

4 a. Give Decision trees for the following set of training examples :

Day	Outlook	Temperature	Humidity	Wind	Play
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

(10 Marks)

b. Discuss practical issues in learning Decision tree. **(10 Marks)**

Module-3

5 a. Explain appropriate problems for neural network learning with its characteristics. **(10 Marks)**

b. Explain single perceptron with its representational power and training rule. **(10 Marks)**

OR

6 a. Explain and derive gradient descent algorithm. **(10 Marks)**

b. Describe the derivation of back propagation rule. **(10 Marks)**

Module-4

7 a. Brief the relation between MAP and ML hypothesis with mathematical notation. Analyse the following statement by deriving appropriate equations :
'Under our choice for P(n) and P(D/n), every consistent hypothesis has posterior probability of $(1 / |VS_{H, D}|)$ and every inconsistent hypothesis has posterior probability of 0, Every consistent hypothesis is therefore a MAP hypothesis. **(10 Marks)**

b. Explain how minimum description length recommends choosing the hypothesis that minimizes the sum of two description lengths. **(10 Marks)**

OR

8 a. We consider a medical diagnosis task. We have knowledge that over the entire population of people, 0.8% have COVID. There exists a binary lab test that represents an imperfect indicator of this disease. The test returns a correct positive result in 98% of the cases in which the disease is present, and a correct negative results in 97% of the cases where the disease is not present.

i) Suppose we observe a patient for whom the lab test returns a positive result. Calculate the posterior probability that this patient truly suffers from COVID.

ii) Knowing that the lab test is an imperfect one, a second independent test is conducted. Calculate the posterior probabilities for COVID and ¬ COVID given that the second test returned a positive result as well. **(10 Marks)**

b. Maximum likelihood hypothesis is the one that minimizes the sum of the squared errors between the observed training values and the hypothesis predications. Justify the above statement with relevant equations. **(10 Marks)**

Module-5

9 a. Differentiate the error rate of the hypothesis over the sample available data and over unknown distribution. How good you can estimate one error provided the another one. **(10 Marks)**

b. Consider the following data set.

0	+
+	0

Identify whether or not the following algorithms can classify this dataset with zero error. Justify answer :

i) KNN with k = 3

ii) Regression. **(05 Marks)**

c. What is curse of dimensionality? How this can be overcome by distance weighted KNN? **(05 Marks)**

OR

10 a. Describe k – nearest neighbor algorithm with its mathematical model and an example. **(10 Marks)**

b. Explain the terminologies used in the phase 'locally weighted regression'. Write your remarks on this algorithm. **(10 Marks)**

* * * * *