

Model Question Paper

Sixth Semester BE Degree Examination

Introduction to Machine Learning

Time: 3 Hours

Max. Marks: 100

Note: 1. Answer any **FIVE** full questions, choosing **ONE** full question from each module.
 2. M: Marks, L: RBT (Revised Bloom's Taxonomy) level, C: Course outcomes.

Module -1			M	L	C																									
Q1	a.	Explain the following a) Concept Learning b) Inductive learning hypothesis c) Consistent Hypothesis	06	L2	CO1																									
	b.	Explain the categories of Machine Learning Algorithms	06	L2	CO1																									
	c.	With an example dataset explain the FIND-S Algorithm to determine the hypothesis	08	L2	CO1																									
OR																														
Q2	a.	Explain inductive bias in candidate elimination algorithm	06	L2	CO1																									
	b.	Explain the following activities of Machine Learning & discuss the significance of these activities in success of machine learning application a) Feature Extraction b) Feature Engineering c) Model Building and Feature Selection	06	L2	CO1																									
	c.	With an example dataset explain the Candidate Elimination Algorithm to determine the hypothesis.	08	L2	CO1																									
Module- 2																														
Q3	a.	Write a python code to any suitable synthetic dataset to mark and remove missing values and illustrate the working of program.	07	L3	CO2																									
	b.	Write a python code to apply following data cleaning methods to example dataset given in table 1 and discuss the need of respective data cleaning process. 1. Identifying columns that contain a single value 2. Remove columns that have a low variance																												
	Table 1 : Example Dataset 1																													
Q3	<table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Feature_1</th><th>Feature_2</th><th>Feature_3</th><th>Feature_4</th><th>Feature_5</th></tr> </thead> <tbody> <tr><td>10</td><td>0.01</td><td>5</td><td>1</td><td>20</td></tr> <tr><td>12</td><td>0.02</td><td>6</td><td>1</td><td>21</td></tr> <tr><td>11</td><td>0.01</td><td>5</td><td>1</td><td>19</td></tr> <tr><td>13</td><td>0.02</td><td>7</td><td>1</td><td>22</td></tr> </tbody> </table>					Feature_1	Feature_2	Feature_3	Feature_4	Feature_5	10	0.01	5	1	20	12	0.02	6	1	21	11	0.01	5	1	19	13	0.02	7	1	22
Feature_1	Feature_2	Feature_3	Feature_4	Feature_5																										
10	0.01	5	1	20																										
12	0.02	6	1	21																										
11	0.01	5	1	19																										
13	0.02	7	1	22																										
c.	Write a python code to apply ordinal encoding approach to any suitable synthetic dataset for encoding categorical attribute and illustrate the working of ordinal encoder	07	L3	CO2																										
OR																														
Q4	a.	Write a python program to apply mutual informative approach for categorical feature selection to any suitable synthetic dataset and illustrate the working of program.	07	L3	CO2																									

b.	Write a python program to apply data standardization to any suitable synthetic dataset and illustrate the need & working of data standardization.	06	L3	CO2
c.	Write a python code to identify and remove near zero-variance predictor and discuss the need of removing near zero variance predictor	07	L3	CO2

Module – 3

Q5	<p>a. A retail company tracks monthly sales (in Rs) and the advertisement budget (in Rs) for five months. Considering the following dataset given below, write a python code to build a simple linear regression model using OLS() method approach and print the estimated parameters of regression model and discuss its significance in regression equation.</p> <p>Table 2 : Example Dataset 2</p> <table border="1"> <thead> <tr> <th>Ad Budget (Rs)</th><th>Sales (Rs)</th></tr> </thead> <tbody> <tr> <td>10</td><td>25</td></tr> <tr> <td>15</td><td>35</td></tr> <tr> <td>20</td><td>45</td></tr> <tr> <td>25</td><td>52</td></tr> <tr> <td>30</td><td>60</td></tr> </tbody> </table>	Ad Budget (Rs)	Sales (Rs)	10	25	15	35	20	45	25	52	30	60	07	L3	CO3						
Ad Budget (Rs)	Sales (Rs)																					
10	25																					
15	35																					
20	45																					
25	52																					
30	60																					
Q5	<p>b. Write the functional form equation of the linear regression model and illustrate how does regression parameters & residual errors are computed and examine & illustrate its significance in regression model accuracy</p> <p>c. Write a python code to apply ridge regression to dataset given in Table 4 and illustrate how ridge regression works.</p> <p>Table 3 : Example Dataset 3</p> <table border="1"> <thead> <tr> <th>Area (Sq Ft)</th> <th>1500</th> <th>1800</th> <th>1200</th> <th>2000</th> <th>1700</th> <th>1600</th> <th>2100</th> <th>1300</th> </tr> </thead> <tbody> <tr> <th>Actual Price</th> <td>250K</td> <td>320K</td> <td>200K</td> <td>350K</td> <td>280K</td> <td>270K</td> <td>400K</td> <td>180K</td> </tr> </tbody> </table>				Area (Sq Ft)	1500	1800	1200	2000	1700	1600	2100	1300	Actual Price	250K	320K	200K	350K	280K	270K	400K	180K
Area (Sq Ft)	1500	1800	1200	2000	1700	1600	2100	1300														
Actual Price	250K	320K	200K	350K	280K	270K	400K	180K														

OR

Q6	<p>a. Write a python code to build simple gradient descent based liner regression dataset given in table 3 and illustrate how gradient descent is used for estimating the regression parameters.</p> <p>b. Write a simple loss function used in supervised learning and analyze its significance in machine learning with an example.</p> <p>c. Write a python code to apply lasso regression to dataset given in table 3 and illustrate how lasso regression works.</p>	07	L3	CO3

Module – 4

Q7	<p>a. Write a python code to apply Logistic Regression to classify the heart disease prediction to dataset given in table 4 and illustrate the working of sigmoid function</p>	06	L3	CO4

Table 4 : Example Dataset 4

Age	Cholesterol	HeartDisease
45	230	0
54	270	1
37	190	0
63	310	1
50	250	1
41	200	0
58	290	1
39	180	0

b. With an example demonstrate the working of pandas `get_dummies()` function for encoding categorical features 08 L3 CO4

c. Write a python code to build a confusion matrix for a simple logistic regression model and illustrate the significance of confusion matrix for model performance analysis 06 L3 CO4

OR

Q8	a. Apply k-nearest neighbor algorithm for synthetic dataset and illustrate the workflow of algorithm. 06 L3 CO4
	b. Apply Bayesian belief network to any suitable example and illustrate the working of joint probability destruction for Bayesian inference. 08 L3 CO4
	c. A patient undergoes a lab test for a rare form of cancer. The following information is available: <ul style="list-style-type: none"> a) The probability that a randomly selected person has the cancer is 0.008. b) The lab test returns a positive result in 98% of cases when the cancer is actually present. c) The lab test returns a negative result in 97% of cases when the cancer is not present. d) Suppose we now observe a new patient for whom the lab test returns a positive result. Apply bayes therom & maximum a posteriori hypothesis to diagnose the patient as having cancer or not? 06 L3 CO4
	Module – 5

Q9	a. Write a python code to apply voting classifier to combine random forest, SVM & linear regression for iris-flower classification and illustrate the workflow of the voting classifier 07 L3 CO5
	b. With an example, illustrate the working of adaboost and its significance of adaboost compare to single base model. 06 L3 CO5
	c. With an example, illustrate the working of Random Forest algorithm. 07 L3 CO5

OR

Q10	a. Write a python code to apply linear SVM classification to iris dataset for iris flower classification and illustrate the working of SVM 07 L3 CO5
	b. With an example, illustrate the working of K-Means Clustering algorithm. 06 L3 CO5
	c. With an example, illustrate the working of SVM Regression. 07 L3 CO5
