questions of a section at one place.

Section – A 3 X 5 = 15 Marks

<table>
<tr><th>No.</th><th>Detail of Question</th><th>Marks</th><th>CO</th><th>BL</th><th>KL</th></tr>
<tr><td>1</td><td>What is Bayes' theorem? Write its two applications in Machine Learning?</td><td>3</td><td>3</td><td>R</td><td>F</td></tr>
<tr><td>2</td><td>Given the confusion matrix, find the Classification Accuracy, Recall, Precision, F-measure.
<table>
<tr><th colspan="2" rowspan="2">Actual</th><th colspan="2">Predicated</th></tr>
<tr><th>Positive</th><th>Negative</th></tr>
<tr><td colspan="2">Positive</td><td>6</td><td>4</td></tr>
<tr><td colspan="2">Negative</td><td>2</td><td>8</td></tr>
</table></td><td>3</td><td>3</td><td>A</td><td>P</td></tr>
<tr><td>3</td><td>Consider the following data where fruits and their corresponding value and price are given.
<table>
<tr><th>Fruit</th><th>Value of Fruit</th><th>Price</th></tr>
<tr><td>Avocado</td><td>1</td><td>50</td></tr>
<tr><td>Pineapple</td><td>2</td><td>110</td></tr>
<tr><td>Apple</td><td>1</td><td>25</td></tr>
<tr><td>Mango</td><td>3</td><td>90</td></tr>
<tr><td>Avocado</td><td>3</td><td>120</td></tr>
</table>
Apply the one hot encoding on the dataset.</td><td>3</td><td>1</td><td>A</td><td>P</td></tr>
</table>

4	What are the differences between Classification and Regression?	3	2	U	C
5	What is Curse of Dimensionality? Discuss the steps of the PCA algorithm.	3	5	U	C

Section – B 5 X 3 = 15 Marks

<table>
<tr><th>No.</th><th>Detail of Question</th><th>Marks</th><th>CO</th><th>BL</th><th>KL</th></tr>
<tr><td>1</td><td>We have data from the questionnaires survey (to ask people opinion) and objective testing with two attributes (acid durability and strength) to classify whether a special paper tissue is good or not. Here are six training samples
<table>
<tr><th>X1 = Acid Durability (seconds)</th><th>X2 = Strength (kg/square meter)</th><th>Y = Classification</th></tr>
<tr><td>1</td><td>4</td><td>Good</td></tr>
<tr><td>7</td><td>7</td><td>Bad</td></tr>
<tr><td>7</td><td>4</td><td>Bad</td></tr>
<tr><td>3</td><td>4</td><td>Good</td></tr>
<tr><td>2</td><td>5</td><td>Good</td></tr>
<tr><td>1</td><td>3</td><td>Bad</td></tr>
</table>
Now the factory produces a new paper tissue that pass laboratory test with X1 = 3 and X2 = 7. Without another expensive survey, can we guess what the classification of this new tissue with the help of KNN algorithm. Solve it for both K=3 and K=5.</td><td>5</td><td>2</td><td>A</td><td>P</td></tr>
</table>

<table>
<tr>
<td>2</td>
<td>Consider the following results obtained on correlation of number of hours spent in driving (X) with the risk of developing acute backache (Y). Compute parameters of a linear regression model.

<table>
<tr><td>Number of hours (X)</td><td>10</td><td>9</td><td>2</td><td>15</td><td>10</td><td>16</td><td>11</td><td>16</td></tr>
<tr><td>Risk Score on a scale of 0-100 (Y)</td><td>95</td><td>80</td><td>10</td><td>50</td><td>45</td><td>98</td><td>38</td><td>93</td></tr>
</table>
a) Find the regression line Y = A.X + B.
b) Use the regression line as a model to estimate the Risk Score on 12 hours of driving.</td>
<td>5</td><td>2</td><td>A</td><td>P</td>
</tr>
<tr>
<td>3</td>
<td>How does a machine learning system work? Explain each phase in detail.</td>
<td>5</td><td>1</td><td>U</td><td>C</td>
</tr>
</table>

Time: 2 Hours **Maximum Marks: 30**

Section- A

Note: Attempt All Three Questions. 3 x 2 = 6 Marks

(I) State the types of Machine Learning.

(II) State the difference supervised learning and unsupervised learning.

(III) What's the difference between Type I and Type II error? Differentiate it with the help of an example.

Section- B

Note: Attempt All Three Questions. 3 x 3 = 9 Marks

(I) Define the terms Hypothesis space. Illustrate with an example.

(II) Given the confusion matrix, find: Classification Accuracy, Recall, Precision, F-measure.

N=165		Predicted		
		yes	no	
Actual	yes	100	5	105
	no	10	50	60
		110	55	

(III) Differentiate between Traditional Programming and Machine Learning. Briefly explain with one example each.

Section – C

Note: Attempt Any Three Questions. 3 x 5 = 15 Marks

(I) With respect to following sample space for events A and B

A holds	T	T	F	F	F	F	T
B holds	T	F	T	F	T	F	F

We have P(A)=4/7, P(B)=3/7, P(B/A) =2/4, P(A/B) =2/3

Verify the correctness of Bayes' Theorem.

(II) It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%.

Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

(III) The values of independent variable x and dependent value y are given below:

x	1	2	3	4	5	6	7
y	9	8	10	12	11	13	14

(III) The values of independent variable x and dependent value y are given below:

x	1	2	3	4	5	6	7
y	9	8	10	12	11	13	14

Find the least square regression line y=ax+b. Estimate the value of y when x is 9.

(IV) Using the following equation for classifier;

$0.5X_1+0.5X_2 \geq 0$ (Class A)

$0.5X_1+0.5X_2 < 0$ (Class B)

Classify the test data $X_1=1$ and $X_2= -2$, $X_1=-2$ and $X_2= 1$

Note: Attempt All Three Questions. 3 x 2 = 6 Marks

I. Write the hypothesis, parameters and cost function of linear regression with multiple variables.

II. Compute the weighted sum of Gini Index for 'Past Trend' using the following dataset for Decision Tree:

Past Trend	Open Interest	Trading Volume	Return
Positive	Low	High	Up
Negative	High	Low	Down
Positive	Low	High	Up
Positive	High	High	Up
Negative	Low	High	Down
Positive	Low	Low	Down
Negative	High	High	Down
Negative	Low	High	Down
Positive	Low	Low	Down
Positive	High	High	Up

III. What is 'Overfitting' in Machine learning? how can you avoid overfitting?

I. What is Gradient Descent? Write the pseudo code of Gradient Descent algorithm. How does learning rate affect Gradient Descent algorithm?

II. What is One Hot Encoding? Consider the following data where fruits and their corresponding value and price are given.

Fruit	Value of Fruit	Price
Apple	1	5
Mango	2	10
Apple	1	15
Orange	3	10

Write the one hot encoding of the dataset.

III. Suppose, you have given the following data where x and y are the 2 input variables and Class is the dependent variable.

X	Y	Class
-1	1	-
0	1	+
0	2	-
1	-1	-
1	0	+
1	2	+
2	2	-
2	3	-

a) Draw the scatter plot which shows the above data in 2D space.

b) Suppose, you want to predict the class of new data point x=1 and y=1 using Euclidean distance in 3-NN. In which class this data point belongs to?

c) Suppose, you want to predict the class of new data point x=1 and y=1 using Euclidean distance in 5-NN. In which class this data point belongs to?

I. How does a machine learning system works? Explain each phase in detail.

II. Write the steps of Find-S algorithm in concept learning. Apply the Find-S algorithm and find out the most specific hypothesis on the following dataset:

The concept of the problem will be on what days does a person likes to go on walk.

Time	Weather	Temperature	Company	Humidity	Wind	Goes
Morning	Sunny	Warm	Yes	Mild	Strong	Yes
Evening	Rainy	Cold	No	Mild	Normal	No
Morning	Sunny	Moderate	Yes	Normal	Normal	Yes
Evening	Sunny	Cold	Yes	High	Strong	Yes

III. What do you understand by a confusion matrix? Consider a two-class classification problem and the following values for the parameters:

$$TP = 30, TN = 930, FP = 30, FN = 10$$

Find out accuracy, precision, recall and F1 score.

IV. The heights and weights of a sample of 11 students are:

Height (m)h	1.36	1.47	1.54	1.56	1.59	1.63	1.66	1.67	1.69	1.74	1.81
Weight (kg) w	52	50	67	62	69	74	59	87	77	73	67

a) Compute the regression line.

b) Use the regression line to estimate the weight of someone whose height is 1.6m.

Q1. How Principal Component analysis algorithms works, explain with formulation.

Q2.Differentiate cross validation and bootstrapping in machine learning.

Q3. Estimate the singular of values of matrix $A = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$

Section – B

Attempt all questions. Each carry equal marks. 3 x 3 = 9 Marks

Q1. Consider the following results obtained on correlation of number of hours spent in driving (X) with the risk of developing acute backache (Y). Derive a linear regression model for this data, and give the values of linear coefficients.

Number of hours (X)	10	9	2	15	10	16	11	16
Risk Score on a scale of 0-100 (Y)	95	80	10	50	45	98	38	93

Q2. Differentiate supervised and un-supervised machine learning.Discuss any two techniques of data imputation with their pros and cons in machine learning.

Q3. Consider the following table of data, representing the marks of 5 students in 3 subjects such as Mathematics (M), English (E), and Art (A). Find the co-variance matrix of this data. Enlist the subject(s) with high variability. Extract two principle components from the dataset.

M	E	A
90	60	90
90	90	30
60	60	60
60	60	90
30	30	30

Q1. How experience (E) is related with performance measure (P) in order to perform a task (T) using machine learning? Enlist the steps to design a learner in machine learning.

Q2.Differentiate over-fitting and under-fitting in machine learning.

Q3. Consider two statistically independent events A and B. Prove that P(AB)=P(A)*P(B) using Bayes rule.

Section – B

Attempt all questions. Each carry equal marks. 3 x 3 = 9 Marks

Q1. Consider the following results obtained on correlation of number of hours spent in driving (X) with the risk of developing acute backache (Y). Derive a linear regression model for this data, and give the values of linear coefficients.

Number of hours (X)	10	9	2	15	10	16	11	16
Risk Score on a scale of 0-100 (Y)	95	80	10	50	45	98	38	93

Q2. Differentiate supervised and un-supervised machine learning. Discuss any two techniques of data imputation with their pros and cons in machine learning.

Q3. Consider the following table of data, representing the marks of 5 students in 3 subjects such as Mathematics (M), English (E), and Art (A). Find the co-variance matrix of this data. Enlist the subject(s) with high variability. Extract two principle components from the dataset.

M	E	A
90	60	90
90	90	30
60	60	60
60	60	90
30	30	30

Q1. Enlist the characteristics of a good and effective machine learning model in terms of bias and variance. Which type of fitting of data ensures high bias and low variance?

Q2. Differentiate hypothesis and hypothesis space. Also discuss about the inductive bias.

Q3. Enlist the steps to design a learner in machine learning.

Section – B

Attempt all questions. Each carry equal marks. 3 x 3 = 9 Marks

Q1. Consider the following Equation(1) representing the relationship between height (H) and weight (W) of the students in a class. Derive the equations to estimate the value of parameters (a,b) using least square estimation method.

$$W=aH+b \quad (1)$$

Q2. Consider the following table of data, representing the marks of 5 students in 3 subjects such as Mathematics (M), English (E), and Art (A). Find the co-variance matrix of this data. Extract two principle components from the dataset.

M	E	A
9	6	9
9	9	3
6	6	6
6	6	9
3	3	3

Q3. Define Machine Learning (defined by Tom Mitchell)? Find the outliers in given array.

A= [2,5,6,3,8,4,10,13,16,17,12,11,24,25]

Section A

(3x2=6)

I. Discuss the Gradient Descent approach for finding the best fit regression line.

II. Given the confusion matrix, find: Classification Accuracy, Recall, Precision, F-measure.

		Predicated	
		Positive	**Negative**
Actual	**Positive**	6	4
	Negative	2	8

III. Find a Hypothesis by Find-S for the following instances in a training set.

Origin	Manufacturer	Color	Decade	Type	Class
Japan	Honda	Blue	1980	Economy	Positive
Japan	Toyota	Green	1970	Sports	Negative
Japan	Toyota	Blue	1990	Economy	Positive
USA	Chrysler	Red	1980	Economy	Negative
Japan	Honda	White	1980	Economy	Positive

Section B

(3x3=9)

I. Define Machine learning. Briefly explain the types of learning with one example each.

II. The values of independent variable x and dependent value y are given below:

X	Y
0	2
1	3
2	5
3	4
4	6

Find the least square regression line y=ax+b. Estimate the value of y when x is 10.

III. What is a Categorical Data? What are the Problems with Categorical Data? Discuss One-Hot Encoding to convert Categorical data to Numerical data.