L.J Institute of Engineering and Technology, Ahmedabad.
Applied Machine Learning Practice book (SEM-VII-2025) T1

Note : This Practic book is only for reference purpose. LJU Test question paper may not be completely set from this practice book.

Sr No	Unit Number	Question_Text	MCQ Answer	Marks	Option A	Option B	Option C	Option D
UNIT 1 - Introduction to Machine Learning								
TOPIC NAME:- Overview of ML and use cases, Types of learning - supervised, unsupervised, and reinforcement, Machine Learning models: Classification, Regression, Clustering, Reinforcement								
MCQ								
1	1	_____ is the application of machine learning.	D	0.5	Sentimental analysis	Traffic prediction	speech and face recognition	All of the metioned
2	1	Which of the following is the main goal of machine learning?	A	0.5	Enable computers to learn data	To automate manual tasks	To make computers intelligent	To generate self-aware machines
3	1	Which of the following ML algorithms can be used with unlabelled data?	C	0.5	Instance-based algorithms	Regression algorithms	Clustering algorithm	Classification algorithm
4	1	_____ is a supervised learning task.	D	0.5	Reinforcement learning	Dimensionality reduction	Clustering	Classification
5	1	Which of the below is an example of an unsupervised learning algorithm?	C	1	Decision Trees	Linear Regression	K-Means Clustering	Support Vector Machines
6	1	If machine learning involves target variable it is called _____ Learning.	A	1	Predictive	Descriptive	Reinforcement	None of the mentioned
7	1	In classification, what does the term "class label" refer to?	B	1	The name of the model	The predicted category of an input	The output of a regression model	The input features of a model
8	1	ML comprises learning algorithms that	D	1	Improve their performance	Over time with experience	At executing some task	All of the mentioned
9	1	What do we call an application of machine learning methods to large databases?	C	1	Big data computing	Artificial intelligence	Data mining	Internet of Things (IoT)
10	1	What is the key difference between supervised and unsupervised learning?	A	1	Supervised learning requires labeled data, while unsupervised learning does not.	Supervised learning predicts labels, while unsupervised learning discovers patterns.	Supervised learning is used for classification, while unsupervised learning is used for regression.	Supervised learning is always more accurate than unsupervised learning.
11	1	Which of the below is not a supervised machine learning algorithm?	A	1	K-means	Naïve Bayes	SVM for classification problems	Decision tree
12	1	In machine learning, what is a model?	B	1	A physical machine that performs computations	A representation of the data	A set of training examples	A type of computer algorithm
13	1	Which type of machine learning algorithm aims to mimic the process of human learning?	C	1	Supervised learning	Unsupervised learning	Reinforcement learning	Deep learning
14	1	Which type of learning is characterized by an agent learning through interactions with an environment and receiving rewards?	C	1	Supervised learning	Unsupervised learning	Reinforcement learning	Deep learning
15	1	Identify the kind of learning algorithm for “facial identities for facial expressions”.	B	1	Prediction	Recognition Patterns	Recognizing Anomalies	Generating Patterns

L.J Institute of Engineering and Technology, Ahmedabad.
Applied Machine Learning Practice book (SEM-VII-2025) T1

Note : This Practic book is only for reference purpose. LJU Test question paper may not be completely set from this practice book.

Sr No	Unit Number	Question_Text	MCQ Answer	Marks	Option A	Option B	Option C	Option D
UNIT 2 - Supervised Learning - Classification-I								
TOPIC NAME:- Classification learning steps, K-nearest Neighbors Support Vector Machines								
MCQ								
16	2	How does the Support Vector Machine (SVM) algorithm work in classification?	A	1	It fits a linear hyperplane to separate classes	It performs clustering based on data density	It uses decision trees to make predictions	It finds the nearest neighbors for each data point
17	2	How does the choice of a kernel impact the performance of a Support Vector Machine (SVM)?	C	1	The kernel has no impact on SVM performance	The kernel determines the SVM's maximum margin	The kernel defines the transformation of input features into a higher-dimensional space	The kernel influences the learning rate during training
18	2	a) Hard SVM is the learning rule in which return a hyperplane that separates the training set with the largest possible margin. b) The output of hard-SVM is the separating hyperplane with the largest margin.	D	1	false false	false true	true false	true true
19	2	In the context of Support Vector Machines (SVM), what is the "kernel trick"?	D	1	A technique to add more features to the dataset	A method to increase the regularization term	way to perform dimensionality reduction	A way to implicitly map data to higher-dimensional spaces
20	2	What is a potential drawback of using a very large value of K in KNN?	B,C	1	It may lead to overfitting.	It may lead to underfitting.	It increases computational complexity.	It reduces the accuracy of the model.
Theory Descriptive: NUMERICALS								
21	2	You have the following dataset with two features (X1, X2) and their corresponding classes: X1-----X2--------Class 1-------2-------A 2-------3-------A 3-------1-------B 6-------5-------B 7-------7-------B You want to classify a new point (4, 4) using KNN with K=3. Calculate the Euclidean distance from the new point to each of the existing points and determine the class of the new point.		3				

L.J Institute of Engineering and Technology, Ahmedabad.
Applied Machine Learning Practice book (SEM-VII-2025) T1

Note : This Practic book is only for reference purpose. LJU Test question paper may not be completely set from this practice book.

Sr No	Unit Number	Question_Text	MCQ Answer	Marks	Option A	Option B	Option C	Option D
22	2	A company wants to predict whether potential customers will purchase a new fitness tracker based on their demographic information and fitness habits. The company has collected data from previous customers, including their age, income, and whether they purchased the fitness tracker. Classify a New Customer:(Age: 29,Income: 65 (in thousands)) With k= 3 Consider the following dataset : Age Income (in thousands) Purchased (Yes/No) 25 50 Yes 30 60 Yes 22 45 No 35 70 Yes 40 80 No 28 55 Yes 50 90 No		3				
23	2	You have a dataset with 100 samples, and you are testing the KNN algorithm with different values of K (1, 3, 5, 10). If the accuracy of the model is as follows: K=1: 90% K=3: 85% K=5: 80% K=10: 75% Discuss how the choice of K affects the model's performance based on the given accuracy values.		3				
24	2	The distances and classes of the 3 nearest neighbors to a new point are as follows: Neighbor 1: Distance = 1, Class = A Neighbor 2: Distance = 2, Class = B Neighbor 3: Distance = 3, Class = A Calculate the weights for each neighbor and determine the predicted class for the new point based on weighted voting. Consider Weight as 1/distance.		2				
		Theory Descriptive: PROGRAMS						
25	2	Consider Social_Netwok_Ads.csv dataset - (UserID, Gender, Age, EstimatedSalary, Purchased). Use Age and EstimatedSalary as input features and Purchased as target feature. Split test data set 30% of complete dataset. Build two models of support vector classifier in python using sklearn library, one for linear and another for RBF kernel with C and gamma parameters set. Predict test labels and print test accuracy.		3				

L.J Institute of Engineering and Technology, Ahmedabad.
Applied Machine Learning Practice book (SEM-VII-2025) T1

Note : This Practic book is only for reference purpose. LJU Test question paper may not be completely set from this practice book.

Sr No	Unit Number	Question_Text	MCQ Answer	Marks	Option A	Option B	Option C	Option D
26	2	Implement an SVC model to classify iris flowers into three species (Setosa, Versicolor, and Virginica) based on their sepal and petal dimensions. The dataset contains 150 samples with four features: sepal length, sepal width, petal length, and petal width. Load the Dataset from sklearn.Split the dataset into training and testing sets.Train the KNN Model using the training data. Visualize the Results		3				
27	2	Develop a Support Vector Classifier to predict whether a tumor is malignant or benign based on 30 features computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. The dataset contains 569 samples with binary labels indicating tumor type. (breast_cancer dataset of sklearn) Load the Dataset from sklearn.Split the dataset into training and testing sets.Train the SVC Model using the training data. Visualize the Results Explain the role of the 'random_state' parameter in train_test_split.		3				

28	2	Identify and correct errors in the following code snippet (provide modified version with annotations):
		python
		
		# Incorrect code
		svc_model = SVC()
		svc_model.fit(X_test, y_test)
		predictions = svc_model.predict(X_train)
				3				

29	2	Write a Python program to implement SVM classification for breast cancer prediction with the following requirements: Data preprocessing using StandardScaler use kernel parameters for tuning Why is feature scaling necessary before applying SVM?		4				
TOPIC NAME:- Naïve Bayes								
			MCQ					
30	2	Which of the following statement is not true about Naïve Bayes classifier algorithm?	A	0.5	It cannot be used for Binary as well as multi-class classifications	It is one of the fast and easy machine learning algorithms to predict a class of test datasets	It performs well in Multi-class prediction as compared to other algorithms	It is the most popular choice for text classification problems
31	2	What is the assumptions of Naïve Bayesian classifier?	C	0.5	It assumes that features of a data are completely dependent on each other	It assumes that each input variable is dependent and the model is not generative	It assumes that each input attributes are independent of each other and the model is generative	It assumes that the data dimensions are dependent and the model is generative

L.J Institute of Engineering and Technology, Ahmedabad.
Applied Machine Learning Practice book (SEM-VII-2025) T1

Note : This Practic book is only for reference purpose. LJU Test question paper may not be completely set from this practice book.

Sr No	Unit Number	Question_Text	MCQ Answer	Marks	Option A	Option B	Option C	Option D
32	2	Arrange the following steps in sequence in order to calculate the probability of an event through Naïve Bayes classifier. I. Find the likelihood probability with each attribute for each class. II. Calculate the prior probability for given class labels. III. Put these values in Bayes formula and calculate posterior probability. VI. See which class has a higher probability, given the input belongs to the higher probability class.	B	1	I, II, III, IV	II, I, III, IV	III, II, I, IV	I, III, II, IV
33	2	Which of the following syntaxt is correct to load iris dataset from sklearn?	C	1	from sklearn.dataset load iris	from sklearn.datasets load_iris	from sklearn.datasets import load_iris	from sklearn.dataset import load iris
34	2	Which of the following syntaxt is correct to call Gaussin Naïve Bayes classifier from sklearn?	A	1	from sklearn.naive_bayes import GaussianNB	from sklearn.bayes import Gaussinbayes	from sklearn.naive_bayes import GaussinNaiveBayes	from sklearn.naivebayes import GaussianNaivebBayes
35	2	There are two boxes. The first box contains 3 white and 2 red balls whereas the second contains 5 white and 4 red balls. A ball is drawn at random from one of the two boxes and is found to be white. Find the probability that the ball was drawn from the second box?	B	1	53/50	50/104	54/104	54/44
		Theory Descriptive: PROGRAMS						
36	2	The breast cancer dataset is available in the sklearn.datasets module and can be loaded using load_breast_cancer(). implement a Naive Bayes Classifier on the Breast Cancer Dataset using python's sklearn library. Assume that all features of datasets are continuous variables and must be used for building model. Perform following task. 1. load data set 2. Split data for training and testing 3. Build a model 4. Predict labels of test data		3				
37	2	Implement a Gaussian Naive Bayes classifier to predict whether a patient has diabetes based on various health metrics of PIMA.csv dataset . The dataset consists of information from 768 female Pima Indians aged 21 and older, initially gathered by the National Institute of Diabetes and Digestive and Kidney Diseases. Target variable: Diabetes (binary, 0 or 1) Attributes: Pregnancies, OGTT (Oral Glucose Tolerance Test), Blood pressure, Skin thickness, Insulin, BMI (Body Mass Index), Age, Pedigree diabetes function. 1. Load the Dataset from a CSV file. Provide summary statistics for the dataset. 2. Split the data into training and testing sets (80% train, 20% test). 3. Instantiate a Gaussian Naive Bayes model and fit it on the training data. 4. Predict diabetes status for the test set. Discuss any biases in the dataset and how they may affect model performance.		4				

L.J Institute of Engineering and Technology, Ahmedabad.
Applied Machine Learning Practice book (SEM-VII-2025) T1

Note : This Practic book is only for reference purpose. LJU Test question paper may not be completely set from this practice book.

Sr No	Unit Number	Question_Text	MCQ Answer	Marks	Option A	Option B	Option C	Option D
38	2	Build a Naive Bayes classifier to analyze the sentiment of movie reviews (positive or negative). The IMDB_Dataset.csv typically contains two main columns: Review: The text of the movie review. Sentiment: A label indicating the sentiment of the review, usually categorized as 'positive' or 'negative'. Load the dataset and inspect the structure. Clean the text data by removing punctuation, converting to lowercase, and tokenizing the reviews.Use CountVectorizer or TfidfVectorizer to convert the text data into a numerical format suitable for model training. Split the dataset and Create a Multinomial Naive Bayes model and fit it on the training data. Predict the sentiment of the test reviews.		5				
39	2	Use sklearn.datasets.fetch_20newsgroups() dataset (all sets) to classify newsgroup documents into their respective categories using a Multinomial Naive Bayes classifier. Load the dataset and explore the categories available. Preprocess the text data using CountVectorizer to convert text documents into a matrix of token counts. Split the dataset and Create a Multinomial Naive Bayes model. Fit the model on the training data and print accuracy.		5				

NUMERICALS

40	2	You have a dataset with 5 emails classified as either Spam or Not Spam, and the occurrence of the words in these emails is given in a table: Classify a new_mail with words “Free” and “Win” using above dataset. What is confidence of classified label?		3				
41	2	Disease Diagnosis Dataset Classify a new patient with Symptom A present, Symptom B absent, and Symptom C absent, and provide the confidence of the classified label.		3				

Contains "free"	Contains "win"	Contains "prize"	Class
Yes	Yes	Yes	Spam
Yes	Yes	No	Spam
Yes	No	No	Not-spam
No	Yes	No	Not-spam
No	Yes	Yes	Spam

Symptom A	Symptom B	Symptom C	Class
Yes	Yes	No	Disease
Yes	No	Yes	Disease
No	Yes	Yes	No Disease
No	No	No	No Disease
Yes	Yes	Yes	Disease
No	Yes	No	No Disease
Yes	No	No	Disease

L.J Institute of Engineering and Technology, Ahmedabad.
Applied Machine Learning Practice book (SEM-VII-2025) T1

Note : This Practic book is only for reference purpose. LJU Test question paper may not be completely set from this practice book.

<table>
<tr><th>Sr No</th><th>Unit Number</th><th>Question_Text</th><th>MCQ Answer</th><th>Marks</th><th>Option A</th><th>Option B</th><th>Option C</th><th>Option D</th></tr>
<tr><td>42</td><td>2</td><td>a customer segmentation dataset with two classes: "Loyal" and "New". We are given the following probabilities:

1. The probability of a customer being Loyal: 0.6
2. The probability of a customer being Loyal given that they are a Frequent Buyer and were referred: 0.85
3. The probability of a customer being a Frequent Buyer and being referred given that they are a New customer: 0.2

What is the probability of a customer being a Frequent Buyer and being referred, given that they are a Loyal customer?</td><td></td><td>4</td><td></td><td></td><td></td><td></td></tr>
<tr><td>43</td><td>2</td><td>Dataset For Naive Bayes Classification is given:

identify the species of an entity with the following attributes.
X={Color=Green, Legs=2, Height=Tall, Smelly=No}</td><td></td><td>3</td><td colspan="4"><table><tr><th>Sl. No.</th><th>Color</th><th>Legs</th><th>Height</th><th>Smelly</th><th>Species</th></tr><tr><td>1</td><td>White</td><td>3</td><td>Short</td><td>Yes</td><td>M</td></tr><tr><td>2</td><td>Green</td><td>2</td><td>Tall</td><td>No</td><td>M</td></tr><tr><td>3</td><td>Green</td><td>3</td><td>Short</td><td>Yes</td><td>M</td></tr><tr><td>4</td><td>White</td><td>3</td><td>Short</td><td>Yes</td><td>M</td></tr><tr><td>5</td><td>Green</td><td>2</td><td>Short</td><td>No</td><td>H</td></tr><tr><td>6</td><td>White</td><td>2</td><td>Tall</td><td>No</td><td>H</td></tr><tr><td>7</td><td>White</td><td>2</td><td>Tall</td><td>No</td><td>H</td></tr><tr><td>8</td><td>White</td><td>2</td><td>Short</td><td>Yes</td><td>H</td></tr></table></td></tr>
<tr><td>44</td><td>2</td><td>Find out whether the object with attribute Confident = Yes, Sick = No will Fail or Pass using Bayesian classification.</td><td></td><td>3</td><td colspan="4"><table><tr><th>Confident</th><th>Studied</th><th>Sick</th><th>Result</th></tr><tr><td>Yes</td><td>No</td><td>No</td><td>Fail</td></tr><tr><td>Yes</td><td>No</td><td>Yes</td><td>Pass</td></tr><tr><td>No</td><td>Yes</td><td>Yes</td><td>Fail</td></tr><tr><td>No</td><td>Yes</td><td>No</td><td>Pass</td></tr><tr><td>Yes</td><td>Yes</td><td>Yes</td><td>Pass</td></tr></table></td></tr>
<tr><td colspan="9">TOPIC NAME:- Logistic Regression</td></tr>
<tr><td colspan="9">MCQ</td></tr>
<tr><td>45</td><td>2</td><td>For input x and output y, given that z=w.x+b and y=f(z), Sigmoid function in logistic regression is?</td><td>B</td><td>1</td><td>1/(1+e^z)</td><td>1/(1+e^(-z))</td><td>1/(1-e^z)</td><td>1/(1-e^(-z))</td></tr>
</table>

L.J Institute of Engineering and Technology, Ahmedabad.
Applied Machine Learning Practice book (SEM-VII-2025) T1

Note : This Practic book is only for reference purpose. LJU Test question paper may not be completely set from this practice book.

Sr No	Unit Number	Question_Text	MCQ Answer	Marks	Option A	Option B	Option C	Option D
46	2	In a logistic regression problem, what is a possible output for a new instance?	A	1	0.85	-0.19	1.2	89%
47	2	The output in a logistic regression problem is yes (equivalent to 1 or true). What is its possible value given by algorithm or method?	D	1	Greater than 0.5	Greater than 0.6	Equal to 1	Depends on the algorithm's threshold value
48	2	Which of the following syntaxt is correct to load xyz.csv file using pandas if 'import pandas as pd' syntax is already called?	A	1	dataset= pd.read_csv('xyz.csv')	dataset= pd.read('xyz.csv')	dataset= pd.csv('xyz.csv')	dataset= pd.read_csv('xyz.csv')
49	2	Which of the following syntaxt is correct to call logistic regression classifier from sklearn?	B	1	from sklearn.naive_bayes import logisticregression	from sklearn.linear_model import LogisticRegression	from sklearn.logisticmodel import logisticregression	from sklearn.logistic_model import LogisticRegression
50	2	Which of the following syntaxt is correct to load xyz.csv file using pandas if import pandas as pd is already called?	A	1	dataset= pd.read_csv('xyz.csv')	dataset= pd.read('xyz.csv')	dataset= pd.csv('xyz.csv')	dataset= pd.read_csv(xyz.csv)
51	2	Which of the following syntaxt is correct to load logistic regression classifier from sklearn?	B	1	from sklearn.naive_bayes import LogisticRegression	from sklearn.linear_model import LogisticRegression	from sklearn.naive_bayes import logisticregression	from sklearn.linear_model import Logisticregression
52	2	The odds ratio is:	B	1	The probability of an event occurring.	The ratio of the odds after a unit change in the predictor to the original odds.	The ratio of the probability of an event happening to the probability of the event not happening.	The ratio of the probability of an event not happening to the probability of the event happening.
		Theory Descriptive: PROGRAMS						
53	2	Imagine a telecommunications company that wants to predict whether a customer will churn (leave the service) based on various features such as age, account length, and monthly charges. The company has historical data on customers, including whether they churned or not. Age: (in years), Account_Length: (in months), Monthly_Charges: (in dollars) Churn: Target variable (1 if the customer churned, 0 otherwise) **'Age': [25, 34, 45, 29, 50, 38, 42, 35, 48, 55],** **'Account_Length': [12, 24, 36, 18, 48, 30, 42, 24, 36, 60],** **'Monthly_Charges': [70, 90, 80, 100, 60, 80, 90, 70, 80, 100],** **'Churn': [0, 1, 0, 1, 0, 1, 0, 1, 0, 1]** build a logistic regression model using python that can predict the probability of a customer churning by spliting 80% of given data for training. Predict targets on trained model of test data. Print accuracy		3				

L.J Institute of Engineering and Technology, Ahmedabad.
Applied Machine Learning Practice book (SEM-VII-2025) T1

Note : This Practic book is only for reference purpose. LJU Test question paper may not be completely set from this practice book.

Sr No	Unit Number	Question_Text	MCQ Answer	Marks	Option A	Option B	Option C	Option D
54	2	build a logistic regression model that can predict whether a customer is likely to churn based on the features in file named customer_churn.csv customer_id: Unique customer ID age: Customer age gender: Customer gender (male/female) account_length: Length of the customer's account (in months) international_plan: Whether the customer has an international plan (yes/no) voice_mail_plan: Whether the customer has a voice mail plan (yes/no) number_vmail_messages: Number of voice mail messages total_day_calls: Total day calls total_night_calls: Total night calls total_intl_calls: Total international calls churn: Whether the customer churned (yes/no)		4				
55	2	Use iris dataset for creating a binary classification problem that predicts whether a flower is of the species "Iris-Virginica" or not. Dataset Features: sepal_length, sepal_width, petal_length, petal_width (in cm) species: Species of the iris flower (Iris-setosa, Iris-versicolor, Iris-virginica). Make Predictions and Print Probabilities with Alter Threshold to 0.6		4				
56	2	Use the Titanic.csv to build a logistic regression model that predicts whether a passenger survived or not. Dataset Features: Pclass: Passenger class (1st, 2nd, or 3rd) Gender: Gender of the passenger (male or female) Age: Age of the passenger (in years) SibSp: Number of siblings or spouses aboard the Titanic Parch: Number of parents or children aboard the Titanic Fare: Ticket fare (in British pounds) Survived: Survival status (0 = No, 1 = Yes) Categorical variable Gender need to be converted to numerical value and Missing values for Age and Fare to befilled with their respective medians. A logistic regression model need to created and trained on the training set. Set a custom threshold to classify the predictions. The test predictions are then evaluated based on this new threshold.		4				

NUMERICALS

L.J Institute of Engineering and Technology, Ahmedabad.
Applied Machine Learning Practice book (SEM-VII-2025) T1

Note : This Practic book is only for reference purpose. LJU Test question paper may not be completely set from this practice book.

Sr No	Unit Number	Question_Text	MCQ Answer	Marks	Option A	Option B	Option C	Option D
57	2	A health researcher is studying the relationship between body mass index (BMI) and the likelihood of developing diabetes. They fit a logistic regression model with the following results: Coefficient for BMI : 0.07 Intercept: -4 Calculate the probability of developing diabetes for a person with a BMI of 30. If the odds of developing diabetes are 5 to 1, what is the corresponding probability?		3				
58	2	A study examines the effect of smoking status (smoker vs. non-smoker) on the likelihood of lung cancer. The logistic regression results are as follows: Coefficient for smoker: 1.2, Intercept : -1 Calculate the odds ratio of developing lung cancer for smokers compared to non-smokers.Interpret the odds ratio in the context of this study.		3				
59	2	A company is interested in predicting the likelihood of a customer purchasing a product based on their age and income. The logistic regression model produced the following coefficients: Coefficient for age : 0.03, Coefficient for income : 0.0001, Intercept: -5. Calculate the probability of purchasing for a 30-year-old customer with an income of $50,000. Discuss how each variable influences the likelihood of purchase.		4				
60	2	Consider data published on n = 27 leukemia patients. The data has a response variable of whether leukemia remission occurred (REMISS=y), which is given by a 1 and 0 in case not occured. The predictor variable is percentage labeling index of the bone marrow leukemia cells (LI=x). Log(odds)= -3.78 + 2.90 LI Find e^(-z) for LI 0.8. If person has LI index as 1.19, find whether remission occurred or not along with probability. What minimum value LI index must have to show leukemia remission occurred (y=1) with confidence of 96%.		4				
61	2	At each of six dose levels, 250 insects are exposed to the substance and the number of insects that die is counted . Using statistical software to calculate the observed probabilities as the number of observed deaths out of 250 for each dose level. The estimated probability the insect dies is =−2.644+0.674X, where X is dose. Predicted probabilities of death (based on the logistic model) for dose 1 to 6.		3				
UNIT 3 - Supervised Learning - Classification -II								
TOPIC NAME:- Working of Decision Tree, Pruning Decision Tree								
MCQ								
62	3	Which of the following is not a Decision tree algorithm?	B	1	CART	DBSCAN	C4.5	ID3
63	3	Which attribute selection measure is used in the ID3 algorithm?	B	1	Chi-square	Information Gain	Entropy	Gini Index
64	3	What is the goal of a decision tree algorithm during training?	C	1	To maximize precision	To minimize error	To minimize impurity	To maximize accuracy

L.J Institute of Engineering and Technology, Ahmedabad.
Applied Machine Learning Practice book (SEM-VII-2025) T1

Note : This Practic book is only for reference purpose. LJU Test question paper may not be completely set from this practice book.

Sr No	Unit Number	Question_Text	MCQ Answer	Marks	Option A	Option B	Option C	Option D
65	3	Which algorithm is used for prune decision trees to avoid overfitting?	D	1	CART	Random Forest	C4.5	Both A and C
		Theory Descriptive: PROGRAMS						
66	3	Write a Python code snippet to create a Decision Tree Classifier using the scikit-learn library. Use a sample dataset (like the Iris dataset) for training and testing.		3				
67	3	Prepare decision tree classifier in python using sklearn library for data set diabetes.csv - (pregnant, glucose, bp, skin, insulin, bmi, pedigree, age, label). Use all features except label as independent variable. Use complete dataset for training and limit the depth of tree upto 3 levels and plot using tree.plotTree()method.		3				
68	3	Using the diabetes dataset, implement a Decision Tree Classifier and determine the importance of each feature in predicting diabetes. Write Python code to visualize the feature importances.		3				
69	3	Implement a Decision Tree Classifier and compare its performance with a SVM Classifier on the iris dataset. Write Python code to display the accuracy of both models.		3				
70	3	Write a Python code create a Decision Tree Classifier using the Iris dataset. Use gini index for generating tree maximum upto 4 levels		3				
71	3	Build dtc for diabetes data set. One with maximum leaf nodes upto 10 and another one with minimum sample per leaf as 5.		3				
72	3	Implement a Decision Tree Classifier and compare its performance with a Logostic Classifier on the diabetes dataset. Write Python code to display the accuracy of both models		4				
73	3	Using the diabetes dataset, implement a Decision Tree Classifier and plot ccp_alphaof built tree against any one DT parameter (node count or maximum depth)		4				
74	3	Using the iris dataset, implement a Decision Tree Classifier and print CCp_alpha of built tree		4				
		NUMERICALS						
75	3	A decision tree classifier is used to predict the failure mode of a boiler system with a dataset of 300 samples, where 200 samples indicate normal operation and 100 samples indicate tube leakage. What is the entropy of the classifier?		2				
76	3	Suppose we have a dataset with 100 examples, 40 of which belong to class A and 60 of which belong to class B. The dataset has one feature, and the feature has two possible values. If 30 examples have the first value of the feature and 70 examples have the second value, what is the information gain?		2				

L.J Institute of Engineering and Technology, Ahmedabad.
Applied Machine Learning Practice book (SEM-VII-2025) T1

Note : This Practic book is only for reference purpose. LJU Test question paper may not be completely set from this practice book.

Sr No	Unit Number	Question_Text	MCQ Answer	Marks	Option A	Option B	Option C	Option D
77	3	A car manufacturer is testing a new engine design that can produce three types of failures: overheating, oil leakage, and poor fuel efficiency. A data has been collected on 500 engines produced by the new design, and has labeled each engine as either failure-free, overheating, oil leakage, or poor fuel efficiency. The data is summarized in the following table: Engine status Number of engines Failure-free 350 Overheating 100 Oil leakage 30 Poor fuel efficiency 20 What is the entropy of the dataset?		2				
78	3	A training dataset collected while navigating autonomous vehicle is shown below. Identify discrete & continuous attributes along with their features. Calculate information gain of two discrete attributes individually		4				
79	3	A government agency wants to investigate the factors that contribute to unsafe braking in cars equipped with an ABS system. They have collected data on vehicle speed, road surface type, brake pedal pressure, and ABS activation type for a set of cars involved in accidents where the ABS system failed to prevent unsafe braking. Built a decision tree based on attribute 'ABS Activation Type' only.		4				
80	3	Given the training dataset, in order to build a decision tree to determine the optimal destination for a drone which attribute amongst altitude, speed, wind, temperature, and weather conditions.is most significant?		4				

Speed	Distance to intersection	Traffic Control Device	Pedestrian Presence	Decision
0.5	0.7	Stop sign	Pedestrian crossing	Slow down
0.2	0.9	Green traffic light	No pedestrian crossing	Stop
0.8	0.4	Red traffic light	No pedestrian crossing	Go
0.6	0.1	Pedestrian crossing	No traffic light	Slow down
0.3	0.5	No traffic sign	Pedestrian crossing	Go
0.9	0.8	No traffic sign	No pedestrian crossing	Stop
0.7	0.2	Green traffic light	No pedestrian crossing	Stop
0.4	0.6	Stop sign	No pedestrian crossing	Slow down
0.1	0.3	No traffic sign	No pedestrian crossing	Stop
0.5	0.5	Red traffic light	Pedestrian crossing	Go

Vehicle Speed (km/h)	Road Surface Type	Brake Pedal Pressure (%)	Activation Type	Label
60	Dry	50	Active	Safe
80	Wet	60	Passive	Unsafe
70	Dry	70	Active	Safe
50	Wet	30	Passive	Unsafe
100	Dry	80	Active	Safe
40	Wet	40	Passive	Unsafe
90	Dry	90	Active	Safe
65	Wet	20	Passive	Unsafe
75	Dry	60	Active	Safe
55	Wet	50	Passive	Unsafe

Altitude	Speed	Wind	Temperature	Weather	Destination
High	Fast	Weak	Warm	Sunny	City
Low	Slow	Strong	Cold	Rainy	Forest
Medium	Medium	Weak	Mild	Cloudy	Beach
High	Slow	Strong	Warm	Cloudy	City
Medium	Fast	Weak	Hot	Sunny	Beach
Low	Medium	Strong	Cold	Rainy	Forest
High	Slow	Weak	Warm	Cloudy	City
Low	Fast	Strong	Hot	Sunny	Forest
Medium	Medium	Weak	Mild	Cloudy	Beach
Low	Slow	Strong	Cold	Rainy	Forest

L.J Institute of Engineering and Technology, Ahmedabad.
Applied Machine Learning Practice book (SEM-VII-2025) T1

Note : This Practic book is only for reference purpose. LJU Test question paper may not be completely set from this practice book.

Sr No	Unit Number	Question_Text	MCQ Answer	Marks	Option A	Option B	Option C	Option D
81	3	At the beginning of an exam, you try to predict whether each problem is easy or difficult (D = + if it is difficult and – if it is easy). Let us assume that you use two observable problem attributes (predicates): • The text length L of the problem (L = 1 if it is long, 0 otherwise) • The amount M of math in the text (M = 1 if there is a lot of math, 0 otherwise) For training data, assume that you have examined 12 previous problems from the homeworks, and have collected the following data: The first line of this table reads as follows: 4 problems for which L = 0 and M = 0 were not difficult (D = −). The second line says: 1 problem for which L = 0 and M = 0 was difficult (D = +). Etc… Note that you observed no problem for which L = 0 and M = 1, or L = 1 and M = 1. Based on this training data, you want to compute a representation of a difficult problem (D) in the form of a decision tree using the two binary attributes L and M. Construct the best decision tree you can for the training data.		5	L M D # 0 0 – 4 0 0 + 1 0 1 – 0 0 1 + 3 1 0 – 1 1 0 + 2 1 1 – 1 1 1 + 0			
TOPIC NAME:- Overfitting Underfitting of Models								
MCQ								
82	3	A model is overfitting when	A	1	train error is low but test error is high	both the train and test errors are high	both the train and test errors are low	train error is high but test error is low
83	3	Assuming that we have a dataset with little noise, a model is underfitting when:	B	1	train error is low but test error is high	both the train and test errors are high	both the train and test errors are low	train error is high but test error is low
84	3	For a fixed training set, by sequentially adding parameters to give more flexibility to the model, we are more likely to observe:	B, D	1	a wider difference between train and test errors	a reduction in the difference between train and test errors	an increased or steady train error	a decrease in the train error
85	3	Polynomial models with a high degree parameter	D	1	always have the best test error (but can be slow to train)	underfit more than linear regression models	get higher training error than lower degree polynomial models	are more likely to overfit than lower degree polynomial models
TOPIC NAME:- Interpreting model: Bias and Variance trade off								
MCQ								
86	3	What does bias in the bias-variance tradeoff represent?	B	1	Overly-simplified Model	Error between average model prediction and ground truth	High error on both test and train data	Average variability in the model prediction for the given dataset

L.J Institute of Engineering and Technology, Ahmedabad.
Applied Machine Learning Practice book (SEM-VII-2025) T1

Note : This Practic book is only for reference purpose. LJU Test question paper may not be completely set from this practice book.

Sr No	Unit Number	Question_Text	MCQ Answer	Marks	Option A	Option B	Option C	Option D
87	3	What does variance in the bias-variance tradeoff represent?	D	1	Overly-complex Model	High error on both test and train data	Error between average model prediction and ground truth	Average variability in the model prediction for the given dataset
88	3	What does high bias indicate in the bias-variance tradeoff?	C	1	Low error on train data and high on test	Overly-complex Model	Overly-simplified Model	Average variability in the model prediction for the given dataset
89	3	What does high variance indicate in the bias-variance tradeoff?	B	1	Error between average model prediction and ground truth	Overly-complex Model	Overly-simplified Model	Low error on train data and high on test
90	3	What is the formula for the error in the bias-variance tradeoff? Error = ?	C	1	bias+variance−irreducible error	bias-variance+irreducibleerror	bias+variance+irreducibleerror	bias-variance-irreducibleerror
		Theory Descriptive: PROGRAMS						
91	3	Write python script to plot training and testing errors of svc model for classification. Assume datset of your choice		2				
92	3	Write python script to plot training and testing errors of decision tree classifier model for classification. Assume datset of your choice		2				