



# ERODE SENGUNTHAR ENGINEERING COLLEGE

(An Autonomous Institution, Affiliated to Anna University)

Thudupathi, Perundurai, Erode - 638057



## III YEAR B.TECH INFORMATION TECHNOLOGY

Credit : 3

### MODEL EXAMINATION - SET 2

Maximum Marks : 100

Course Code : 19IT601

Date : 14.05.2025

Course Name : MACHINE LEARNING

Duration : 3h

Q.No	Part A ( 10 x 1 = 10 Marks )				Bloom's Level	COs
1	Identify the kind of learning algorithm for “facial identities for facial expressions”.				UN	CO1
	a) Prediction	b) Recognition Patterns	c) Recognizing Anomalies	d) Generating Patterns		
2	Identify the type of learning in which labeled training data is used.				UN	CO1
	a) Semi Unsupervised Learning	b) Supervised Learning	c) Reinforcement Learning	d) Unsupervised Learning		
3	Among the following options identify the one which is false regression. ?				RE	CO2
	a) It is used for prediction	b) It is used for Interpretation	c) It relates input to output	d) It discovers Casual Relationship		
4	Analysis of ML algorithm needs _____.				RE	CO2
	a) Statistical Learning Theory	b) Computational Learning Theory	c) Both a & b	d) Does not require any theory		
5	Why is the Viterbi algorithm used in HMMs?				EV	CO3
	a) To estimate the transition probabilities	b) To determine the most probable sequence of hidden states	c) To calculate the likelihood of an observed sequence	d) To train a deep learning model		
6	You are comparing two models with similar performance on a validation set. What is the next step in model selection?				EV	CO3
	a) Choose the model with the simplest architecture	b) Test both models on a separate test set to evaluate their true performance	c) Select the model with the highest number of parameters	d) Ignore the test set and choose the model based on training performance		
7	When would you use Ridge Regression instead of Linear Regression?				AP	CO4

	a) When the target variable is categorical	b) When there are many multicollinear features in the dataset	c) When the model needs to be interpretable	d) When you want to use non-linear relationships between features and target		
8	You have two regression models: Model A with higher training accuracy but lower validation accuracy than Model B. Which model is preferable and why?				EV	CO4
	a) Model A, because training accuracy is more important	b) Model B, because it generalizes better to unseen data	c) Model A, because validation accuracy is not relevant	d) Both are equally good as long as they make predictions		
9	You have applied K-means clustering to a dataset and received 4 clusters. Which of the following metrics would be most useful for evaluating the quality of the clusters?				AP	CO5
	a) The F1-score	b) The silhouette score	c) The root mean squared error	d) The mean squared error		
10	Which of the following methods is commonly used for discovering latent factors in a recommendation system?				UN	CO5
	a) K-means clustering	b) Principal Component Analysis (PCA)	c) Singular Value Decomposition (SVD)	d) Naive Bayes classification		
<b>Q.No</b>	<b>Part B ( 10 x 2 = 20 Marks )</b>				<b>Bloom's Level</b>	<b>COs</b>
11	Differentiate supervised and unsupervised machine learning.				UN	CO1
12	Give the significance of 'Training set' and 'Testing Set'.				UN	CO1
13	Analyse why we need to use Regression?				AN	CO2
14	Analyse Why do we need a K-NN Algorithm?				AN	CO2
15	In spam detection, a Naïve Bayes classifier assigns a high probability to an email being spam. What action should be taken?				AP	CO3
16	A speech recognition system using an HMM struggles to correctly classify words. What could improve its performance?				EV	CO3
17	You are using linear regression on a dataset and notice that the residuals are not randomly distributed. What could this imply about the model's fit?				AN	CO4
18	You observe that your model performs well on the training data but poorly on the test data. What steps can you take to reduce overfitting?				AP	CO4
19	Suggest a modification to improve clustering performance when the dataset contains noise and outliers.				CR	CO5

20	If a recommendation system with latent factors overfits the training data, what strategies would you apply to improve generalization?		CR	CO5																				
Q.No	Part C ( 5 x 14 = 70 Marks )	Marks	Bloom's Level	COs																				
21	a) List all the steps involved in designing a learning system. Apply the same in Checkers Learning Problem and with respect to Tasks, Performance and Experience.	14	AP	CO1																				
	or																							
	b) Explain the significance of Simple Linear Regression. Give its Mathematical Equation. How do we calculate the Mean Squared Error for LR. Consider Sales of Marketing Company A of last 5years	14	AP	CO1																				
<table><tr><td>Year</td><td>Advertisement in expenses</td><td>sales</td></tr><tr><td>2017</td><td>90</td><td>1000</td></tr><tr><td>2018</td><td>120</td><td>1300</td></tr><tr><td>2019</td><td>150</td><td>1800</td></tr><tr><td>2020</td><td>100</td><td>1200</td></tr><tr><td>2021</td><td>130</td><td>1380</td></tr><tr><td>2022</td><td>200</td><td>?</td></tr></table>	Year				Advertisement in expenses	sales	2017	90	1000	2018	120	1300	2019	150	1800	2020	100	1200	2021	130	1380	2022	200	?
Year	Advertisement in expenses				sales																			
2017	90				1000																			
2018	120				1300																			
2019	150				1800																			
2020	100				1200																			
2021	130	1380																						
2022	200	?																						
Compute the prediction using Simple Linear Regression for the year 2022, when the expenses for the Advertisement are 200\$.																								

22	a) Enumerate the Working Principle of KNN algorithm. Derive KNN algorithm. Assume a dataset that contains multiple user's information regarding their purchasing through Online Shopping. The dataset contains lots of information but the Estimated Salary and Age is considered as the independent variables and the Purchased variable is for the dependent variable.	14	AP	CO2																																																		
	<table><tr><td>User ID</td><td>Gender</td><td>age</td><td>Estimated salary</td><td>purchased</td></tr><tr><td>15265410</td><td>male</td><td>25</td><td>82000</td><td>0</td></tr><tr><td>15265411</td><td>male</td><td>32</td><td>52000</td><td>0</td></tr><tr><td>15265412</td><td>male</td><td>45</td><td>19000</td><td>0</td></tr><tr><td>15265413</td><td>male</td><td>46</td><td>15000</td><td>0</td></tr><tr><td>15265414</td><td>male</td><td>48</td><td>14000</td><td>0</td></tr><tr><td>15265415</td><td>female</td><td>25</td><td>12000</td><td>0</td></tr><tr><td>15265416</td><td>female</td><td>27</td><td>18000</td><td>1</td></tr><tr><td>15265417</td><td>female</td><td>28</td><td>16000</td><td>0</td></tr><tr><td>15265418</td><td>female</td><td>29</td><td>17000</td><td>1</td></tr></table>				User ID	Gender	age	Estimated salary	purchased	15265410	male	25	82000	0	15265411	male	32	52000	0	15265412	male	45	19000	0	15265413	male	46	15000	0	15265414	male	48	14000	0	15265415	female	25	12000	0	15265416	female	27	18000	1	15265417	female	28	16000	0	15265418	female	29	17000	1
	User ID				Gender	age	Estimated salary	purchased																																														
	15265410				male	25	82000	0																																														
	15265411				male	32	52000	0																																														
	15265412				male	45	19000	0																																														
	15265413				male	46	15000	0																																														
	15265414				male	48	14000	0																																														
	15265415				female	25	12000	0																																														
	15265416				female	27	18000	1																																														
15265417	female	28	16000	0																																																		
15265418	female	29	17000	1																																																		
or																																																						
b) Define KNN algorithm. How does it Works? Consider we have height, weight and T-shirt size of some customers and we need to predict the T-shirt size of a new customer given only height and weight information we have. Data including height, weight and T-shirt size information is shown below	14	AP	CO2																																																			
<table><tr><td>height (cm)</td><td>weight(kg)</td><td>Tshirt size</td><td>height (cm)</td><td>weight (kg)</td><td>T shirt size</td></tr><tr><td>158</td><td>58</td><td>M</td><td>165</td><td>45</td><td>L</td></tr><tr><td>158</td><td>59</td><td>M</td><td>165</td><td>52</td><td>L</td></tr><tr><td>160</td><td>61</td><td>M</td><td>165</td><td>65</td><td>L</td></tr><tr><td>160</td><td>52</td><td>M</td><td>165</td><td>55</td><td>L</td></tr><tr><td>163</td><td>54</td><td>M</td><td>165</td><td>55</td><td>L</td></tr><tr><td>163</td><td>55</td><td>M</td><td>165</td><td>59</td><td>L</td></tr></table>				height (cm)	weight(kg)	Tshirt size	height (cm)	weight (kg)	T shirt size	158	58	M	165	45	L	158	59	M	165	52	L	160	61	M	165	65	L	160	52	M	165	55	L	163	54	M	165	55	L	163	55	M	165	59	L									
height (cm)				weight(kg)	Tshirt size	height (cm)	weight (kg)	T shirt size																																														
158				58	M	165	45	L																																														
158				59	M	165	52	L																																														
160				61	M	165	65	L																																														
160				52	M	165	55	L																																														
163	54	M	165	55	L																																																	
163	55	M	165	59	L																																																	

23	a) Design an HMM for a complex task, such as recognizing actions in video sequences. What states, observations, and transition/emission probabilities would you define? Explain how you would train this model and validate its performance.	14	CR	CO3
	<b>or</b>			
	b) Provide an example of how HMMs can be applied in a real-world scenario, such as speech recognition or time-series analysis. Demonstrate how HMMs handle sequential data with spatial or temporal dependencies.	14	AP	CO3
24	a) Evaluate model selection strategies and explain the criteria for selecting the right model.	14	EV	CO4
	<b>or</b>			
	b) After applying regularization techniques like L2 regularization and reducing the number of latent factors, you still face issues with generalization. What additional strategies could you use to address overfitting and improve the performance of your recommendation system?	14	AP	CO4
25	a) Design a method to detect fraudulent behavior in a financial transaction network using graph-based approaches. Justify your design using graph properties such as centrality or connectivity.	14	CR	CO5
	<b>or</b>			
	b) A retail chain wants to segment its stores based on sales performance. You are provided with the following dataset, where each point represents a store with two features: X: Average monthly footfall (in thousands) Y: Average monthly revenue (in lakhs) Data points: Store 1: (1, 4), Store 2: (1, 3), Store 3: (5, 8), Store 4: (6, 9), Store 5: (7, 5), Store 6: (8, 6) The company decides to use $k = 2$ for clustering and chooses these initial centroids: Centroid A: (1, 4) Centroid B: (7, 5) find out i) Assign each store to the nearest centroid using Euclidean distance. ii) Calculate the new centroids after the first iteration of k-means	14	AP	CO5

Course Outcomes		Marks
CO1	Explore the acquired knowledge on recalling the applications of machine learning.	34
CO2	Understand the concepts behind different types of learning and their appropriateness.	34
CO3	Choose and apply appropriate learning technique for a given real world problem.	34
CO4	Analyze the observations for a given set of data.	34
CO5	Evaluate the effectiveness of different learning techniques for different kinds of data and applications	34

<b>Blooms Level</b>	UN	RE	AN	AP	EV	CR	<b>Total</b>
<b>Marks</b>	7	2	6	104	19	32	<b>170</b>

<b>Prepared By</b>	<b>Scrutinized By</b>	<b>Verified By</b>
SIVASANKARI S (AP / IT) [ 12-05-2025 ]	Dr.THIRUVENKATASURESH M (Prof. / IT) [ 13-05-2025 ]	Dr.THIRUVENKATASURESH M (Prof. / IT) [ 13-05-2025 ]