

K. J. Somaiya Institute of Engineering and Information Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

End Semester Exam

Nov-Dec 2021

Program: M.Tech. (AI)

Examination: FY Semester: I

Course Code: **1PCEC102** and Course Name: **Machine Learning and Pattern Recognition**

Duration: 03 Hours

Max. Marks: 60

Instructions:

- (1) All questions are compulsory.
- (2) Draw neat diagrams wherever applicable.
- (3) Assume suitable data, if necessary.

		Max. Marks	CO	BT level
Q1	Solve any six questions out of eight:	12		
i)	Define Hypothesis space and inductive bias	2	CO1	U
ii)	What do you mean by curse of dimensionality?	2	CO2	U
iii)	What are the advantages of C4.5 over ID3 classification algorithm?	2	CO3	U
iv)	Why KNN is a lazy learning and non-parametric algorithm.	2	CO4	U
v)	What do you mean by Parameter tuning?	2	CO6	U
vi)	Define conditional probability. What is Bayes' theorem.	2	CO4	U
vii)	What do you mean by Ensemble learning?	2	CO5	U
viii)	What is R-squared in Linear regression? Explain with example.	2	CO3	U
Q.2	Solve any four questions out of six.	16		
i)	Compare with example overfitting, underfitting & best fitting model.	4	CO3	U
ii)	What is the need of data cleaning and data transformation, explain with suitable examples?	4	CO1	U

iii)	Differentiate between feature selection and feature extraction.	4	CO2	U																																																																																										
iv)	Write advantages of Random Forest over Decision tree algorithm.	4	CO5	U																																																																																										
v)	Suppose the data for clustering is {2, 4, 10, 12, 3, 20, 30, 11, 25,5,36,41,14}. Assuming number of clusters to be 2 i.e. $K = 2$, cluster the given data using k-means clustering algorithm.	4	CO4	Ap																																																																																										
vi)	Illustrate Linear regression in machine learning. What are the measures uses for cost function in linear regression algorithm.	4	CO3	U																																																																																										
Q.3	Solve any two questions out of three.	16																																																																																												
i)	Define Ensemble Learning. How Random Forest machine leaning algorithm work. Explain with example.	8	CO5	U																																																																																										
ii)	Perform Linear discriminant analysis for the given 2-D dataset <table border="1" data-bbox="232 950 995 1022"> <tr> <td>$C1 \rightarrow X1=(X1, X2)$</td> <td>(4,2)</td> <td>(2,4)</td> <td>(2,3)</td> <td>(3,6)</td> <td>(4,4)</td> </tr> <tr> <td>$C2 \rightarrow X2=(X1, X2)$</td> <td>(9,10)</td> <td>(6,8)</td> <td>(9,5)</td> <td>(8,7)</td> <td>(10,8)</td> </tr> </table>	$C1 \rightarrow X1=(X1, X2)$	(4,2)	(2,4)	(2,3)	(3,6)	(4,4)	$C2 \rightarrow X2=(X1, X2)$	(9,10)	(6,8)	(9,5)	(8,7)	(10,8)	8	CO2	Ap																																																																														
$C1 \rightarrow X1=(X1, X2)$	(4,2)	(2,4)	(2,3)	(3,6)	(4,4)																																																																																									
$C2 \rightarrow X2=(X1, X2)$	(9,10)	(6,8)	(9,5)	(8,7)	(10,8)																																																																																									
iii)	What is Hyperplane and Support Vectors in the SVM algorithm? What is the use of SVM kernel?	8	CO4	U																																																																																										
Q.4	Solve any two questions out of three.	16																																																																																												
i)	Using the given training dataset classify the above tuples using Naïve Bayes algorithm: (Outlook = sunny, Temperature = cool, Humidity=high and Wind =strong) <table border="1" data-bbox="329 1400 1011 1860"> <thead> <tr> <th>Day</th> <th>Outlook</th> <th>Temperature</th> <th>Humidity</th> <th>Wind</th> <th>PlayTennis</th> </tr> </thead> <tbody> <tr><td>D1</td><td>Sunny</td><td>Hot</td><td>High</td><td>Weak</td><td>No</td></tr> <tr><td>D2</td><td>Sunny</td><td>Hot</td><td>High</td><td>Strong</td><td>No</td></tr> <tr><td>D3</td><td>Overcast</td><td>Hot</td><td>High</td><td>Weak</td><td>Yes</td></tr> <tr><td>D4</td><td>Rain</td><td>Mild</td><td>High</td><td>Weak</td><td>Yes</td></tr> <tr><td>D5</td><td>Rain</td><td>Cool</td><td>Normal</td><td>Weak</td><td>Yes</td></tr> <tr><td>D6</td><td>Rain</td><td>Cool</td><td>Normal</td><td>Strong</td><td>No</td></tr> <tr><td>D7</td><td>Overcast</td><td>Cool</td><td>Normal</td><td>Strong</td><td>Yes</td></tr> <tr><td>D8</td><td>Sunny</td><td>Mild</td><td>High</td><td>Weak</td><td>No</td></tr> <tr><td>D9</td><td>Sunny</td><td>Cool</td><td>Normal</td><td>Weak</td><td>Yes</td></tr> <tr><td>D10</td><td>Rain</td><td>Mild</td><td>Normal</td><td>Weak</td><td>Yes</td></tr> <tr><td>D11</td><td>Sunny</td><td>Mild</td><td>Normal</td><td>Strong</td><td>Yes</td></tr> <tr><td>D12</td><td>Overcast</td><td>Mild</td><td>High</td><td>Strong</td><td>Yes</td></tr> <tr><td>D13</td><td>Overcast</td><td>Hot</td><td>Normal</td><td>Weak</td><td>Yes</td></tr> <tr><td>D14</td><td>Rain</td><td>Mild</td><td>High</td><td>Strong</td><td>No</td></tr> </tbody> </table>	Day	Outlook	Temperature	Humidity	Wind	PlayTennis	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	8	CO4	Ap
Day	Outlook	Temperature	Humidity	Wind	PlayTennis																																																																																									
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																																																																																									

ii)	What do you mean by Hypothesis space and inductive bias in machine learning. Also explain Model visualization?	8	CO1	U																																																																																										
iii)	<p>Apply CART (decision tree) algorithm for the following table. Solve upto first level of tree generation ie. Rootnode node.</p> <table border="1" data-bbox="263 500 1069 991"> <thead> <tr> <th><i>RID</i></th> <th><i>age</i></th> <th><i>income</i></th> <th><i>student</i></th> <th><i>credit_rating</i></th> <th><i>Class: buys_computer</i></th> </tr> </thead> <tbody> <tr><td>1</td><td>youth</td><td>high</td><td>no</td><td>fair</td><td>no</td></tr> <tr><td>2</td><td>youth</td><td>high</td><td>no</td><td>excellent</td><td>no</td></tr> <tr><td>3</td><td>middle_aged</td><td>high</td><td>no</td><td>fair</td><td>yes</td></tr> <tr><td>4</td><td>senior</td><td>medium</td><td>no</td><td>fair</td><td>yes</td></tr> <tr><td>5</td><td>senior</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr> <tr><td>6</td><td>senior</td><td>low</td><td>yes</td><td>excellent</td><td>no</td></tr> <tr><td>7</td><td>middle_aged</td><td>low</td><td>yes</td><td>excellent</td><td>yes</td></tr> <tr><td>8</td><td>youth</td><td>medium</td><td>no</td><td>fair</td><td>no</td></tr> <tr><td>9</td><td>youth</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr> <tr><td>10</td><td>senior</td><td>medium</td><td>yes</td><td>fair</td><td>yes</td></tr> <tr><td>11</td><td>youth</td><td>medium</td><td>yes</td><td>excellent</td><td>yes</td></tr> <tr><td>12</td><td>middle_aged</td><td>medium</td><td>no</td><td>excellent</td><td>yes</td></tr> <tr><td>13</td><td>middle_aged</td><td>high</td><td>yes</td><td>fair</td><td>yes</td></tr> <tr><td>14</td><td>senior</td><td>medium</td><td>no</td><td>excellent</td><td>no</td></tr> </tbody> </table> <p>Table 6.1 Class-labeled training tuples from AllElectronics customer database.</p>	<i>RID</i>	<i>age</i>	<i>income</i>	<i>student</i>	<i>credit_rating</i>	<i>Class: buys_computer</i>	1	youth	high	no	fair	no	2	youth	high	no	excellent	no	3	middle_aged	high	no	fair	yes	4	senior	medium	no	fair	yes	5	senior	low	yes	fair	yes	6	senior	low	yes	excellent	no	7	middle_aged	low	yes	excellent	yes	8	youth	medium	no	fair	no	9	youth	low	yes	fair	yes	10	senior	medium	yes	fair	yes	11	youth	medium	yes	excellent	yes	12	middle_aged	medium	no	excellent	yes	13	middle_aged	high	yes	fair	yes	14	senior	medium	no	excellent	no	8	CO3	Ap
<i>RID</i>	<i>age</i>	<i>income</i>	<i>student</i>	<i>credit_rating</i>	<i>Class: buys_computer</i>																																																																																									
1	youth	high	no	fair	no																																																																																									
2	youth	high	no	excellent	no																																																																																									
3	middle_aged	high	no	fair	yes																																																																																									
4	senior	medium	no	fair	yes																																																																																									
5	senior	low	yes	fair	yes																																																																																									
6	senior	low	yes	excellent	no																																																																																									
7	middle_aged	low	yes	excellent	yes																																																																																									
8	youth	medium	no	fair	no																																																																																									
9	youth	low	yes	fair	yes																																																																																									
10	senior	medium	yes	fair	yes																																																																																									
11	youth	medium	yes	excellent	yes																																																																																									
12	middle_aged	medium	no	excellent	yes																																																																																									
13	middle_aged	high	yes	fair	yes																																																																																									
14	senior	medium	no	excellent	no																																																																																									