PVP 20

Code: 20CS4501A

III B.Tech - I Semester – Regular / Supplementary Examinations
NOVEMBER 2024

DATA SCIENCE
(COMPUTER SCIENCE & ENGINEERING)

Duration: 3 hours Max. Marks: 70

Note: 1. This paper contains questions from 5 units of Syllabus. Each unit carries 14 marks and have an internal choice of Questions.
2. All parts of Question must be answered in one place.

BL – Blooms Level CO – Course Outcome

			BL	CO	Max. Marks
UNIT-I					
1	a)	"The initial investigation of data is done by Data Exploratory". Explain the above statement with the help of summary statistics and graphical representation.	L2	CO1	10 M
	b)	Write a short note on Data Acquisition.	L2	CO1	4 M
OR					
2	a)	Describe in detail about hyper parameter optimization techniques in Data Science.	L2	CO1	10 M
	b)	Discuss briefly on Concept of Deployment in Data Science.	L2	CO1	4 M
UNIT-II					
3	a)	Illustrate Data cleaning techniques and benefits of cleaning the Data.	L3	CO2	7 M

	b)	Compare the various stages of Data Transformation in preprocessing model of Data Science.	L3	CO2	7 M
OR					
4	a)	What does a PCA do? How is the first principal component axis selected?	L3	CO2	7 M
	b)	Discuss in detail about Data reduction with an example.	L3	CO2	7 M
UNIT-III					
5	a)	What is the difference between probability density function (PDF) and probability mass function (PMF)? Write down the proportion they must satisfy.	L2	CO3	10 M
	b)	Write a Short note on Simple Random sampling.	L2	CO3	4 M
OR					
6	a)	Explain in detail about Chi-squared distribution with example.	L2	CO3	10 M
	b)	What is Systematic sampling?	L2	CO3	4 M
UNIT-IV					
7	a)	Explain Logistic Regression with an example.	L3	CO4	7 M
	b)	Which specific regressors seem essential in multiple regressions? How will you address this question? Discuss.	L3	CO4	7 M
OR					

<table>
<tr><td>8</td><td>a)</td><td>Sam found how many hours of sunshine vs how many ice creams were sold at the shop from Monday to Friday:
<table><tr><th>"x" Hours of Sunshine</th><th>"y" Ice Creams Sold</th></tr><tr><td>2</td><td>4</td></tr><tr><td>3</td><td>5</td></tr><tr><td>5</td><td>7</td></tr><tr><td>7</td><td>10</td></tr><tr><td>9</td><td>15</td></tr></table>Let us find the best m (slope) and b (y-intercept) that suits that data using Linear Regression.</td><td>L3</td><td>CO4</td><td>10 M</td></tr>
<tr><td></td><td>b)</td><td>Discuss in short about Least Square Method in Linear Regression</td><td>L3</td><td>CO4</td><td>4 M</td></tr>
<tr><td colspan="6">UNIT-V</td></tr>
<tr><td>9</td><td>a)</td><td>How does model complexity affect bias and variance? What is the difference between Bias and Variance?</td><td>L3</td><td>CO4</td><td>6 M</td></tr>
<tr><td></td><td>b)</td><td>Explain in detail about Bias Variance decomposition for model assessment with example algorithm.</td><td>L3</td><td>CO4</td><td>8 M</td></tr>
<tr><td colspan="6">OR</td></tr>
<tr><td>10</td><td>a)</td><td>Briefly explain how you would calculate a cross-validated estimate of prediction error in a Linear regression. Is this estimate likely more minor or more significant than the in-sample error?</td><td>L3</td><td>CO4</td><td>7 M</td></tr>
<tr><td></td><td>b)</td><td>What is the holdout approach? What is the limitation of this approach? Name four alternative approaches for it.</td><td>L3</td><td>CO4</td><td>7 M</td></tr>
</table>