ME Comp / Sem II / Choice Base / Second Half 2018

Q.P. Code: 40828

19/11/2018

Time: 03 Hours

Marks: 80

Note: 1. Question 1 is compulsory

2. Answer any three out of remaining questions.

3. Assume suitable data wherever required and justify the same.

Q1	a)	Describe how logistic regression can be used as a classifier.	[5]
	b)	Explain the ethical issues in data science.	[5]
	c)	State differences between LDA and PCA.	[5]
	d)	What is sharding? Explain the advantages of sharding.	[5]
Q2	a)	Describe the six steps of data science process.	[10]
	b)	Describe the working of the Map-Reduce with an example.	[10]
Q3	a)	Explain Gaussian (normal) distribution with respect to pdf and cdf and its use in statistics.	[10]

	b)	Given S= 0 0
		       -1 -1
		       1 2
		Find principal components.
			[10]

Q4	a)	What is type I and type II errors in hypothesis testing? Is one always more serious than the other? Why?	[10]
	b)	Explain the process of content based RS with suitable example.	[10]
Q5	a)	Explain singular value decomposition (SVD) with an example.	[10]
	b)	What is No SQL? Compare No SQL with SQL.	[10]
Q6	a)	Explain Independent Component Analysis (ICA) in descriptive modeling.	[10]
	b)	What is gradient descent algorithm and how it helps in minimizing the cost function.	[10]

* * *