Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A constituent unit of MAHE, Manipal)

VII SEMESTER B.TECH END SEMESTER EXAMINATIONS, NOV 2019

SUBJECT: PROGRAM ELECTIVE - V

FUNDAMENTAL TECHNIQUES OF DATA SCIENCE [CRA 4011]

REVISED CREDIT SYSTEM
(26/11/2019)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

- Answer **ALL** the questions.
- Missing data if any, may be suitably assumed.

1A. What do you mean by multiple comparison problem? Explain the various types of errors that occur when one considers a set of statistical inferences simultaneously. Explain the various measures to control the error rate. **5**

1B. What is power? Mention the different factors that affect power. **3**

1C. Write the nonparametric bootstrap procedure for calculating confidence interval for the median from a data set of n observations. **2**

2A. What do you mean by residuals? Explain the properties of residuals. Also, show how to obtain residual variance. **5**

2B. Suppose you take a sample of stock returns from Company_A and Company_B from the years 2008 to 2012 as shown in Table Q.2B. Calculate the empirical covariance and correlation between the stock returns.

Table Q.2B

Year	Company_A	Company_B
2008	1	3
2009	-2	2
2010	3	4
2011	0	6
2012	3	0

3

2C. Define the following with its mathematical equations or examples:

i) Random variable

ii) Survival function **2**

3A. Explain the rules a function must satisfy in order to be a valid probability density function. Let X be a random variable with pdf given by:

$$f_X(x) = \begin{cases} cx^2, & |x| \le 1 \\ 0, & otherwise \end{cases}$$

Find the following: **5**

i) The constant c.
ii) $P(X) \geq 1/2$

3B. Consider the belief network shown in Figure Q.3B. Determine the probability that a fish came from North Atlantic, given that it is spring time and that the fish is light salmon.

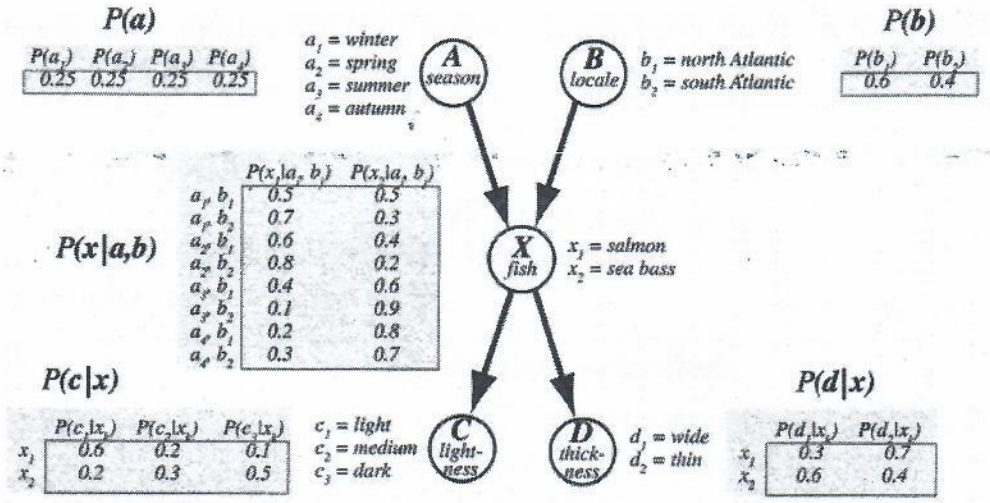

Figure Q.3B

3C. Your boss (to be fair) randomly assigns everyone an extra 2 hours work on weekend evenings between 4 PM and midnight. What are the chances that you get on Saturday between 4PM and 6PM?

4A. Explain expected value of a random variable. Given X is a discrete random variable. The table Q.4A defines a probability distribution for X. What is the expected value of X?

Table Q.4A

x	$P(X = x)$
−7	0.2
−3	0.1
3	0.4
7	0.3

4B. What is the Central Limit Theorem? Explain with an example.

4C. Compare and contrast the following terminologies with mathematical equations:
i) Population mean vs sample mean
ii) Population variance vs sample variance

5A. State and explain the following:
i) Bernoulli distribution
ii) Normal distribution

5B. Find the median, lower quartile, upper quartile, interquartile range and range of the following numbers.
12, 5, 22, 30, 7, 36, 14, 42, 15, 53, 25, 65

5C. What is adjustment with respect to regression models? Explain with an example.