cct

| Reg.No. | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|

# MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

*(A constituent unit of MAHE, Manipal)*

**V SEMESTER B.TECH. (INFORMATION TECHNOLOGY / COMPUTER AND COMMUNICATION ENGINEERING)**

**END SEMESTER EXAMINATIONS, NOVEMBER 2018**

SUBJECT: PROGRAM ELECTIVE I – PATTERN RECOGNITION [ICT 4020]

**REVISED CREDIT SYSTEM**

**(30/11/2018)**

Time: 3 Hours MAX. MARKS: 50

**Instructions to Candidates:**

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

**1A.** Estimate p(θ/D) of the univariate Gaussian distribution, where $\theta = (\hat{\mu}, \hat{\sigma}^2)$, using Bayesian parameter estimation technique and state the significance of the estimated value. 5

**1B.** Compute the Type_of_candidate for the tuple (40, 600000, 27000) using K-nearest neighbor algorithm with k =3 for the data given in Table Q.1B. 3

**Table Q.1B**

| | | | | | | |
|---|---|---|---|---|---|---|
| Age (Years) | 19 | 19 | 27 | 25 | 24 | 37 |
| Savings (Rs.) | 50000 | 30000 | 200000 | 150000 | 8500 | 500000 |
| Monthly_Expenses (Rs.) | 8000 | 6000 | 20000 | 35000 | 6000 | 40000 |
| Type_of_candidate | Student | Student | Professional | Professional | Student | Professional |

**1C.** Prove that the uncorrelated variables need not be independent using relevant example. 2

**2A.** Extract the productions from transmission and emission matrix given in Table Q.2A. Considering S1 as the starting state, check whether the productions yields the strings "eat", "team" and "meet". If yes, compute the total probability of generating the strings. 5

**Table Q.2A**

| Transition Matrix | | Present State | |
|---|---|---|---|
| | | S1 S2 | S2 S3 |
| Previous State | S1 | 0.8 | 0.2 |
| | S2 | 0 | 1 |
| | S3 | 1 | 0 |

| Emission Matrix | m | e | a | t |
|---|---|---|---|---|
| S1 | 0.2 | 0.5 | 0 | 0.5 |
| S2 | 0 | 0.3 | 0.7 | 0 |
| S3 | 0 | 0 | 0.25 | 0.75 |

**2B.** Compute the distance of point (v = (66, 640, 44)) from the data points given in Table Q.2B using Chebyshev distance metric. Assume h =1000. Also, state the properties of distance metrics. 3

**Table Q.2B.**

| | | | | | |
|---|---|---|---|---|---|
| X (Height) | 64 | 66 | 68 | 69 | 73 |
| Y (Score) | 580 | 570 | 590 | 660 | 600 |
| Z (Age) | 29 | 33 | 37 | 46 | 55 |

**2C.** State the "No Free Lunch Theorem" and "Ugly Duckling Theorem". 2

**3A.** Draw the HMM for the data given in Table Q.3A and decode the states emitting sequence 'E1 E2 E1 E1 E3' using Viterbi algorithm. Assume that the probability of states at any random time instance is P(S1) = 0.12, P(S2) = 0.22 and P(S3) = 0.66 **5**

**Table Q.3A**

| Transition Matrix | | Present State | | |
|---|---|---|---|---|
| | | S1 | S2 | S3 |
| Previous State | S1 | 0 | 0.2 | 0.8 |
| | S2 | 0.5 | 0 | 0.5 |
| | S3 | 0 | 0.3 | 0.7 |

| Emission Matrix | E1 | E2 | E3 |
|---|---|---|---|
| S1 | 0.6 | 0.4 | 0 |
| S2 | 0.3 | 0 | 0.7 |
| S3 | 0 | 0.8 | 0.2 |

**3B.** Consider the Table. Q.3B representing marks distribution for a two class category problem. Identify the category of given marks 66 using Bayes decision rule. Assume the prior probability of obtaining first class or second class is 0.5 each. **3**

**Table Q.3B**

| Marks | 62 | 10 | 36 | 45 | 78 | 89 | 25 | 95 | 59 | 97 |
|---|---|---|---|---|---|---|---|---|---|---|
| Class | First | Second | Second | Second | First | First | Second | First | Second | First |

**3C.** Outline the limitations of Bayesian decision making. **2**

**4A.** Differentiate between the following: **5**

i) PCA and LDA
ii) Parametric and Nonparametric approach
iii) HMM and BBNs
iv) Bagging and boosting techniques
v) Parzen window and KNN

**4B.** Explain the techniques to overcome missing and noisy features in the data set used for pattern classification. **3**

**4C.** Bayes decision rule is used for minimizing probability of error. Justify this statement using suitable graph and equations. **2**

**5A.** Obtain the Sum of Squared Error (SSE) and the new values of 'a' and 'b' after first iteration, using gradient descent for a linear regression algorithm. Consider the data given in Table Q.5A. Assume the initial value of a = 0.1, b = 0.2, r =0.01. Use min-max normalization technique to normalize the values. **5**

**Table Q.5A.**

| Quantity in kg (X) | 11 | 9 | 2 | 5 | 6 | 4 |
|---|---|---|---|---|---|---|
| Price (Y) | 300 | 250 | 50 | 130 | 150 | 100 |

**5B.** Explain the importance of feature extraction and classifier evaluation during design cycle of pattern recognition system. **3**

**5C.** Calculate the Jackknife estimate of bias and variance of mode using histogram plot for **2**

$$D = \{15, 15, 20, 20, 20, 30, 30\}.$$