Department of Computer Science and Engineering, MNIT Jaipur
CST-905 Research methodology

Time: 2 hrs 30 min **Max. Marks: 30**

Attempt **any six** questions. All questions carry equal marks. Show your work so that partial marks can be awarded even if the final result is not correct,

1) A distance metric [*d*(*x, y*) : distance between points *x* and *y*] should satisfy the following conditions:
   d(*x, x*) = 0
   d(*x, y*) = *d*(*y, x*)
   d(*x, z*) <= *d*(*x, y*) + *d*(*y, z*)
   Verify if Euclidean distance in 2D and *n*-dimensional space meet above conditions.

2) Covariance between two variables X and Y is defined as
$$\text{cov}(x, y) = \frac{\sum_{k=1}^{N} (X_k - \mu_k)(Y_k - \mu_k}{N - 1}$$
   Using above definition, prove the following (any two)
   (i) cov(*aX*, *Y*)=*a* cov(*Y, X*)
   (ii) cov(*X*+*a*, *Y*) = cov(*Y, X*)
   (iii) cov(*X*+*Y*, *Z*) = cov(*X, Z*) + cov(*Y, Z*)

3) Two batteries are randomly chosen from a group of 3 new, 2 used but still working, and 2 defective batteries. If we let X denote the number of new and Y denote used but still working batteries that are chosen, then determine the joint probability mass function of *X* and *Y* , *p*(*i,* j) = *P*{*X*=*i, Y*=*j*}.

OR

   p(*x, y*), the joint probability mass function of X and Y, is as follows
   p(0, 0) = 0.30,
   p(0, 1) = 0.25,
   p(1, 0) = 0.10,
   p(1, 1) = 0.35.
   Calculate the conditional probability mass function of X given that (a)Y=0, (b) Y = 1.

4) What are various sampling methods to select a representative sample from a given population? List shortcomings of each method. Describe in your own words why random sampling is a better choice.

5) The Joint density function of X and Y is given by
$$f(x, y) = 2e^{-2x}e^{-y} \qquad x \geq 0, \quad y \geq 0$$
   Compute
   (a) P(X > 1, Y > 1}
   (b) P(X > Y}
   (c) P(X < a}, a is non-negative
   (d) P(Y > a}, a is non-negative

6) (a) Write short notes on any two data visualisation techniques along with the limitations of each.
   (b) Draw box-plot of the following data. X = {1,3,2,4,5,7,2,1,6,5,2,3,4,8,7,3,2,4,1,8}

7) What is machine learning? Explain in brief each of the following ML techniques highlighting merits and limitations of any two of the following
   (i) Linear perceptron
   (ii) Binary Decision Diagram
   (iii) Neural Network
   (iv) Support Vector Machine

8) Differentiate between supervised and unsupervised learning. What does training mean in context of Machine Learning? Discuss (i) underfittng and (ii) overfitting in training a model? Explain why should both of these lead to a bad learning model.