Birla Institute of Technology & Science, Pilani (Pilani Campus)
CS F320 Foundations of Data Science

Comprehensive Examination

Part-A (Closed book)

Date: 15/05/2023 Weightage: 17.5% Duration: 90 minutes Max. marks: 35

Instructions:

I. Please write all the answers in the Part-A answer sheet only. Do not submit this question paper.

II. For True/False, and fill in the blanks, you only need to write your answer in the answer sheet.

1. Mark the following statements **True** or **False**: [11]

   i). When the variance of the model is high, overfitting is less.

   ii). When the feature space is large, overfitting is less.

   iii). As the number of training examples increases, the model trained on that data will have lower variance.

   iv). For high-dimensional objects, most of their volume is near the surface.

   v). Images/videos are examples of structured data.

   vi). Maximum likelihood estimation (MLE) is based on the frequentist approach.

   vii). Parameters estimates using a linear support vector machine (SVM) has higher variance than non-linear SVM (with Gaussian kernel) on the same dataset.

   viii). The optimal solution for the optimization problem in SVM is a globally optimal solution.

   ix). In the case of strong duality, the primal and dual solutions are different.

   x). The time-series generated from a random walk is not stationary.

   xi). The time-series of IID noise has trend/seasonality component.

2. Fill in the blanks: [6]

   i). A valid kernel function satisfies the _____ condition.

   ii). The splitting criteria for a decision tree can be based on _____ impurity.

   iii). A decision tree with high depth has _____ variance.

   iv). White noise is an example of _____ (stationary/ non-stationary) process.

   v). As the number of dimensions increases, the volume of a cube _____ (increases/decreases/remains constant).

   vi). In comparison to non-linear SVM, linear SVM has _____ bias, and _____ variance.

3. State the primal problem of SVM. Derive the dual problem of SVM using the Lagrange multiplier method showing all the steps. [2+4=6]

4. List the expressions for the forward and backward propagation in a 2 layer neural network. [3]

5. Find the VC dimension of threshold function and an interval. [3]

6. List out three problems which arise in the analysis of time series data. [3]

7. Write all the steps involved in the dynamic time warping algorithm. Show the minimum cost path for the alignment between two time-series, X=[0, 2, 0], and Y=[0, 0, 0.5]. [2+1=3]

———— * * * * * * * * * ————

Birla Institute of Technology & Science, Pilani (Pilani Campus)

CS F320 Foundations of Data Science

Comprehensive Examination

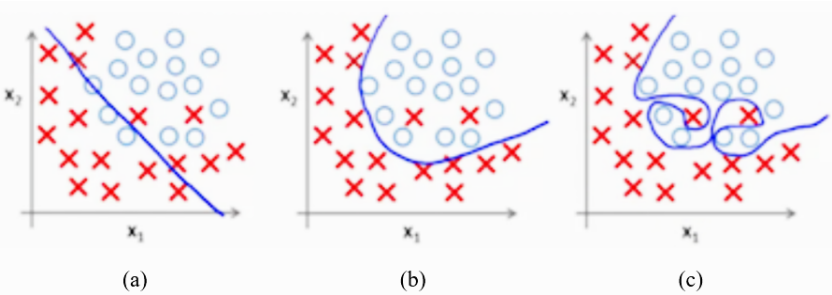
Part-B (Open book)

Date: 15/05/2023 Weightage: 17.5% Duration: 90 minutes Max. marks: 35

1. Can the following Boolean function be represented by a single unit of a neural network? If so, find the weights and bias associated with the unit. Also, show it pictorially in a graph. [3]

A	B	f(A,B)
1	1	0
0	0	0
1	0	1
0	1	0

2. Explain each of the following scenarios with respect to the fitting of the classifier. [3]



3. a). Can the XOR function (see table below) be classified using logistic regression? Justify your answer. [2]

   b). Design a neural network using logistic units (as in logistic regression) to classify the XOR function ($A.\overline{B} + \overline{A}.B$) as shown below. Also, show the network pictorially. [5]

A	B	f(A,B)
0	0	0
0	1	1
1	0	1
0	1	0

4. Calculate the principal components for the following data showing all the steps of the PCA. Also, show it in a graph. [4]

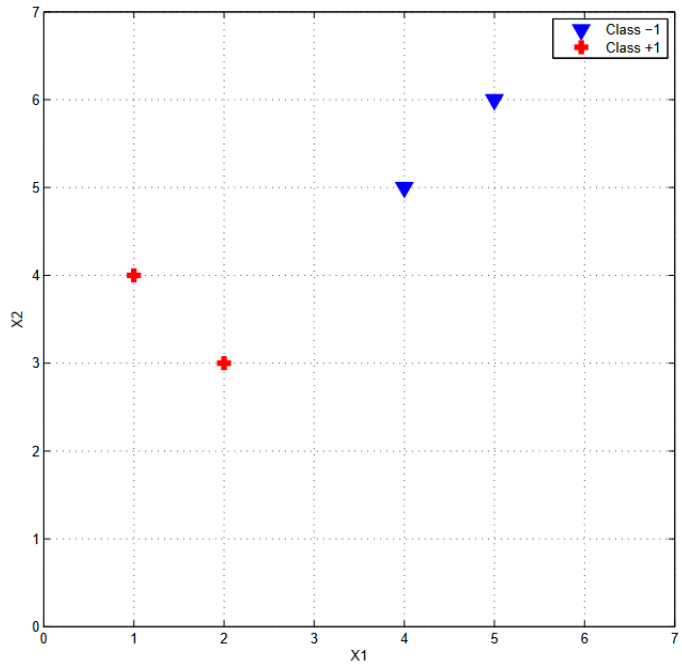
X	Y
1	-1
0	1
-1	0

5. Find the maximum and minimum values of the function

   a). $f(x, y) = xy$, subject to the constraint $4x^2 + y^2 = 8$. [4]

   b). $f(x, y) = x^2 + 2y^2$, subject to the constraint $x^2 + y^2 \leq 4$ [4]

6. Consider the following dataset shown in the plot. [2+1+1=4]



a). Find the maximal margin decision boundary for SVM in this case, by using system of equations and geometrical interpretation of this dataset.

b). List the support vectors, weight vector and bias

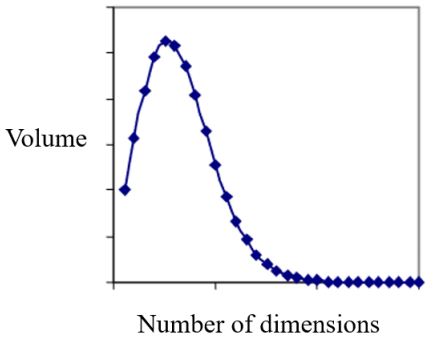
c). Show the decision boundary and support vectors in the plot.

7. A model of social mobility of families posits three different social classes (strata), namely 'lower', 'middle', and 'upper'. The transitions between these classes (states) for a given family are governed by the following transition matrix

$$A = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 2/3 \end{bmatrix}$$

Find the stationary distribution of this Markov chain. [3]

8. Consider the plot below



a). This plot is a representation of which geometrical object?

b). Write the formula for the volume of the object?

c). Give proper justification for the change in volume with increasing dimension in this plot. [1+1+1=3]

———— * * * * * * * * * ————