Register No:
Date:

ST. JOSEPH'S COLLEGE (AUTONOMOUS), BANGALORE-27
M.SC. BIG DATA ANALYTICS - II SEMESTER
SEMESTER EXAMINATION: APRIL 2018
BDA 2316: MACHINE LEARNING I

TIME: 1 ½ HRS **MAX MARKS 35**

This Question Paper Contains TWO Printed Pages
Answer any five questions of the following
All questions carry equal marks (7 marks)

1 Linear Regression

Suppose that a dataset contains p explanatory variables $X_1, X_2, \ldots, X_p$ and one response variable Y. The standard linear model is defined as follows:

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \cdots + \theta_p X_p$$

Briefly describe the *Gradient Descent* procedure that may be used to find the parameters $\theta_0, \theta_1, \theta_2, \ldots, \theta_p$ based on the given dataset as training data.

2 Logistic Regression

Suppose that given a dataset with p explanatory variables $X_1, X_2, \ldots, X_p$ and one binary response variable Y (can assume values 0 or 1), your logistic regression model for classifying the cases $Y = 0$ and $Y = 1$ is as follows:

$$\Pr(Y = 1) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 X_1 + \theta_2 X_2 + \cdots + \theta_p X_p)}}; \quad \Pr(Y = 0) = 1 - \Pr(Y = 1).$$

Would you decide $Y = 1$ if $\Pr(Y = 1 \mid X_1, X_2, \ldots, X_p) \geq 0.5$, or would you go for a different decision threshold, something other than 0.5? What is the effect of changing this decision threshold on the accuracy of classification?

3 Neural Networks

What is the relation between a logistic regression model for classification and a single perceptron model with a sigmoid/logistic activation function? How is a multi-layer neural network able to form non-linear decision boundaries? Briefly explain the principle behind *backpropagation* algorithm in case of a multi-layer neural network. How is gradient descent used in neural networks?

4 Support Vector Machines

What is a *support vector* classifier? What is the role of the support vectors? Briefly explain the advantage of using a non-linear kernel in a support vector machine, and name a few popular non-linear kernels used in practical SVMs.

5 Unsupervised Learning

Suppose you are given a p-dimensional dataset with n points, and you can't view the whole dataset in a plot as the dimension p is quite high. Briefly describe the k-means algorithm to form k clusters in the n-point dataset. Briefly justify how you would judge if your choice of k is *right* in this case.

6 Dimensionality Reduction

Suppose you are given a p-dimensional dataset with n points, and you can't view the whole dataset in a p-dimensional plot as the value of p is too high. Briefly describe a method to *meaningfully* reduce the dimension of the dataset to $d = 2$ so that it may be viewed easily. Is the total *information* in the d-dimensional reduced dataset the same as the original p-dimensional dataset? Can you quantify the fraction of information still preserved?

7 Anomaly Detection

How would you identify if a data point is regular or is an anomaly? Briefly explain one strategy to perform anomaly detection in high-dimensional data. Explain the challenges of anomaly detection in case of a time-series dataset.

8 System Design

Explain the phenomenon of *bias-variance trade-off* in case of modelling. Given a training dataset consisting of n observations, briefly describe one method of *cross-validation* to ensure robust predictions on the test dataset. Briefly describe an approach to compare two given models for classification.