Date:

Registration number:



ST. JOSEPH’S COLLEGE (AUTONOMOUS), BENGALURU-27

MSc COMPUTER SCIENCE II SEMESTER

SEMESTER EXAMINATION: APRIL 2022

(Examination conducted in July 2022)

**CS 8221 – Machine Learning With Python**

Time- 2 ½ hrs Max Marks-70

This question paper contains FOUR printed pages and THREE parts

**Part A**

**Answer all questions 1X15=15**

1. Machine learning is a field of AI consisting of learning algorithms that excel at-----------

a) Executing some task

b)Learning

c)Learning over time withb experience.

d)All of the above

1. What is called the average squared difference between classifier predicted output and actual output?

a)Mean relative error

b)Mean squared error

c)Mean absolute error

d)Root mean squared error.

1. How can you handle missing or corrupted data in dataset?

a)Drop missing rows or columns

b) Assign a unique category to missing values

c)  Replace missing values with mean/median/mode

d)All of the above.

1. Machine Learning algorithms build a model based on sample data known as -----------------

a)Transfer data

b)Training data

c) Data training

d) None of the above.

1. A machine learning technique that helps in detecting outliers in the data is called -------------------

a)Clustering

b)Classification

c)Anomaly detection

d) All of the above

1. In the following type of feature selection we start with empty feature set

a) Forward elimination

b)Forward Selection

c) Backward elimination

d) All of the above.

1. PCA can be used for ------------------------

a)Projection and visualization of data in higher dimensions

b) Visualization of data

c) projection and visualization of data in lower dimensions

d)None of the above

1. Of the Following Examples, Which would you address using an supervised learning Algorithm?

a) given email labeled as spam or not spam, learn a spam filter

b) given a set of news articles found on the web, group them into set of articles about the same story.

c)given a database of customer data, automatically discover market segments and group customers into different market segments.

d) find the patterns in market basket analysis

## 9. Which of the following is a good test dataset characteristic?

a) large enough to yield meaningful results

b) is representative of the dataset as a whole

c) both a and b

d). none of the above

10. Support Vector Machine is

a)logical model

b)proababilistic model

c) geometric model

d) none of the above

## 11.  A perceptron adds up all the weighted inputs it receives, and if it exceeds a certain value, it outputs a 1, otherwise it just outputs a 0.

a) true

b) false

c) sometimes – it can also output intermediate values as well

d) can’t say

## 12.What is the purpose of the Kernel Trick?

a) to transform the data from nonlinearly separable to linearly separable

b)to transform the problem from regression to classification

c) to transform the problem from supervised to unsupervised learning.

d) all of the above

13.The selling price of a house depends on many factors. For example, it depends on the number of bedrooms, number of kitchen, number of bathrooms, the year the house was built, and the square footage of the lot. Given these factors, predicting the selling price of the house is an example of \_\_\_\_\_\_\_\_\_\_\_\_ task.

a)binary classification

b) multilabel classification

c) simple linear regression

d)multiple linear regression

## 14. What are tree based classifiers?

## a)classifiers which form a tree with each attribute at one level

b)classifiers which perform series of condition checking with one attributeat a time

c)both a and b

d)none of the options

## 15. What does K refers in the K-Means algorithm which is a non-hierarchical clustering approach?

a) complexity

b)fixed value

c)no of iterations

d)number of clusters

**Part B**

**Answer five questions 5x5=25**

16. A)Explain the differences between supervised and un supervised learning.

OR

B)What is underfitting and overfitting ?What are its importance in machine learning?

17. A)Explain linear regression model with a suitable example.

OR

B)What is dimensionality reduction ? How principal component analysis support dimensionality reduction?

18. A)What is Bayes’s theorem and maximum posterior hypothesis?

OR

B)Explain working of back propagation algorithm for multilayer feed forward network.

19.A)State and explain the principles of Hidden markov model.

OR

B) Explain the characteristics of reinforcement learning .

20.A) Explain the steps in python how to apply k means algorithm on the given data for

k=3 .Use c1=2,c2=16,c3=38 as initial cluster centres.

Given dataset 2,4,6,3,31,38,15,16,12 .

B) Write in detail about the different data plots which are available in MATPLOTLIB.

**Part C**

**Answer any three questions 3x10=30**

21.Explain the terms Gain and Entropy .How is it used to build decision tree in ID3 algorithm. Illustrate using an example.

22.The following table gives data set about stolen vehicles .Using Naive bayes classifier classify the new data(Red ,SUV,Domestic)

|  |  |  |  |
| --- | --- | --- | --- |
| **Color** | **Type** | **Origin** | **Stolen** |
| Red | Sports | Domestic | Yes |
| Red | Sports | Domestic | No |
| Red | Sports | Domestic | Yes |
| Yellow | Sports | Domestic | No |
| Yellow | Sports | Imported | Yes |
| Yellow | SUV | Imported | No |
| Yellow | SUV | Imported | Yes |
| Yellow | SUV | Domestic | No |
| Red | SUV | Imported | No |
| Red | Sports | Imported | Yes |

23. What is the goal of Support Vector Machine (SVM)?Explain How to compute margin SVM algorithm.

24. Write short notes on the following

a)Types of activation functions (4 marks)

b) Learning rate (2 marks)

c) Parametric and non parametric models(4 marks)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*