SRM Institute of Science and Technology

School of Computing

DEPARTMENT OF DATA SCIENCE AND BUSINESS SYSTEMS

SRM Nagar, Kattankulathur – 603203, Chengalpattu District, Tamilnadu

Academic Year: 2024-25 (ODD)

Test: CLAT-1 [Portion: UNIT 1 & First half of 2nd Unit]	**Date: 23.02.2024**
Course Code & Title: 21CSE426T & Financial Machine Learning	**Duration:** 100 Minutes
Year & Sem: III Year & VI Sem	**Max. Marks:** 40

Part A [Answer all / 10 * 1 = 10 Marks]

1. Dimensionality reduction is the process of reducing the ____________ in a dataset while preserving information and overall model performance.
   a. Features b. coefficients c. Hyperparameters d. relations
2. Finding the best combination of hyperparameters of a model: ___________
   a. RMSE b. Model tuning c. Search d. SSE
3. Finding the best combination of hyperparameters of a model: ___________
   a. RMSE b. Model tuning c. Search d. SSE
4. The number of hidden nodes in each layer of ANN is ___________ of input layer & output layer size
   a. sum b. difference c. mean d. square
5. The cost function of OLS is ____________
   a. RSS b. RMSE c. SSE d. MSE
6. The dataset to be used in Linear Regression should not have : ____________
   a. multicollinearity b. multilinearity c. multigrid d. multisearch
7. $CostFunction = RSS + \lambda * \sum_{j=1}^{p} |\beta_j|$ This is the formula that represents ___________ regularization:
   a. Lasso b. Rigid c. Elastic net d. Elastic grid
8. _____________the coefficients lead to a lower variance and a lower error value.
   a. Zeroing b. Shrinking c. Expanding d. Changing
9. logistic regression models the ____________________ of the output classes
   a.values b. labels c. Probabilitiesd. coefficients
10. Using kernels, the original data is projected into a _________ to classify the data better.
    a.Lower dimension b. Higher dimension c. hyperplane d. boundaries

Part B [Answer any 2 / 3 * 15= 30 marks]

1. You are tasked with designing a simple ANN to predict whether a patient has a particular disease based on various health metrics (e.g., blood pressure, cholesterol level, age). Explain the following:
   - The architecture of the ANN, including the choice of activation functions and the number of layers.
   - Write a Python code snippet using **TensorFlow/Keras** to create and compile this ANN model.
   - Describe the steps involved in training this model on a dataset and how you would evaluate its performance.
2. You are provided with a dataset containing information about different types of fruits (e.g., apples, oranges, bananas). You need to classify the fruits based on their physical characteristics such as weight, color, and size.
   - Explain the concept of the kernel trick in SVM and its importance in non-linear classification.
   - Write a Python code snippet using **sklearn** to implement the classifier with an RBF kernel for this task.
   - Outline the steps to perform data preprocessing, feature scaling, and model evaluation.
3. You are working on a binary classification problem where you need to predict whether a customer will buy a product based on their demographic information (e.g., age, income, location).
   - Explain the concept of the logistic function and its role in logistic regression.
   - Write a Python code snippet using **sklearn** to implement logistic regression for this classification
   - Discuss the steps involved in feature selection, model training, and interpreting the output coefficients.