

**20IT7301: DEEP LEARNING**

**HOME ASSIGNMENT-1 QUESTIONS**

A.Y:2023-24

|  |  |  |  |
| --- | --- | --- | --- |
| Batch  No | QUESTION | CO | BTL |
|  | **COMMON for all QUESTIONS:**  Link to download the dataset corresponding to the task: https://drive.google.com/drive/folders/1Y5stBLqJpB3uFICmDsXYfuMasK30fo5F?usp=s hare\_link  **Check Task number and regression/classification task. Based on that download the dataset.**   1. Build an ANN model for the given real world problem. 2. Consider the required hyper parameter values of your own. 3. Analyze the performance of your model with two different epoch values and show the loss and accuracy comparison tables and discuss. | CO1,CO4 | APPLY, analyze |
| 1 | TASK 1: CLASSIFICATION: STROKE PREDICTION According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths. This dataset is used to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status. Each row in the data provides relevant information about the patient. | CO1,CO4 | APPLY, analyze |
| 2 | TASK 1:  REGRESSION: WIND SPEED PREDICTION High precision and reliable wind speed forecasting is a challenge for meteorologists. Severe wind due to convective storms, causes considerable damages (large scale forest damage, outage, buildings/houses damage, etc.). Convective events such as thunderstorms, tornadoes as well as large hail, strong winds, are natural hazards that have the potential to disrupt daily life, especially over complex terrain favoring the initiation of convection. Even ordinary convective events produce severe winds which causes fatal and costly damages. Therefore, wind speed prediction is an important task to get advanced severe weather warning. This dataset contains the responses of a weather sensor that collected different weather variables such as temperatures and precipitation.  Find output image by considering input and filter matrices given | CO1,CO4 | APPLY, analyze |
| 3 | Task2  CLASSIFICATION: HEART FAILURE PREDICTION Cardiovascular diseases (CVDs) are the number 1 cause of death globally, CO 1-4 Apply, Analyze and taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide. Four out of 5CVD deaths are due to heart attacks and strokes, and one-third of these deaths occur prematurely in people under 70 years of age. Heart failure is a common event caused by CVDs and this dataset contains 11 features that can be used to predict a possible heart disease. People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidaemia or already established disease) need early detection and management wherein a machine learning model can be of great help. | CO1,CO4 | APPLY, analyze |
| 4 | Task2  REGRESSION: FUEL CONSUMPTION RATINGS Dataset provides model-specific fuel consumption ratings and estimated carbon dioxide emissions for new light-duty vehicles for retail sale in Canada in 2022. | CO1,CO4 | APPLY, analyze |
| 5 | TASK 3: CLASSIFICATION: HEPATITIS C PREDICTION An infection caused by a virus that attacks the liver and leads to inflammation. The virus is spread by contact with contaminated blood; for example, from sharing needles or from unsterile tattoo equipment. Most people have no symptoms. Those who do develop symptoms may have fatigue, nausea, loss of appetite and yellowing of the eyes and skin. Hepatitis C is treated with antiviral medication. In some people, newer medicines can eradicate the virus. Hence your task predicting the HEPATITIS C with given features. | CO1,CO4 | APPLY, analyze |
| 6 | Task3 : REGRESSION: DAILY GOLD PRICE PREDICTION Historically gold coinage was widely used as currency; when paper money was introduced, it typically was a receipt redeemable for gold coin or bullion. In a monetary system known as the gold standard, a certain weight of gold was given the name of a unit of currency. For a long period, the United States government set the value of the US dollar so that one troy ounce was equal to $20.67 ($0.665 per gram), but in 1934 the dollar was devalued to $35.00 per troy ounce ($0.889/g). By 1961, it was becoming hard to maintain this price, and a pool of US and European banks agreed to manipulate the market to prevent further currency devaluation against increased gold demand. | CO1,CO4 | APPLY, analyze |
| 7 | TASK 4: CLASSIFICATION: PREDICTING THE ABSENTEEISM AT WORK Workplace absences always have a personal component to them. This can make it hard for a company to discern the particular cause without having an employee divulge information about his personal life. In specific cases, that may eventually become necessary, but in general companies can fight absenteeism by making it more appealing for their employees to come to work and by showing understanding and lenience when absences occur. In particular, companies should strive to provide their workers with enough pay, enough dignity and a positive enough workplace so that merely showing up each day doesn’t wear a worker down. These are all major factors affecting absenteeism. Absenteeism costs a lot in terms of bottom lines and productivity, according to AIHR Digital. | CO1,CO4 | APPLY, analyze |
| 8 | Task4 : REGRESSION: PREDICT SALES OR PROFIT AT SUPERMARKET STORE With growing demands and cut-throat competitions in the market, a Superstore Giant is seeking your knowledge in understanding what works best for them. They would like to understand which products, regions, categories and customer segments they should target or avoid. You can even take this a step further and try and build a Regression model to predict Sales or Profit. Go crazy with the dataset, but also make sure to provide some business insights to improve. | CO1,CO4 | APPLY, analyze |
| 9 | TASK 5: CLASSIFICATION: IN-VEHICLE COUPON RECOMMENDATION Coupon systems have been widely used to enhance customers’ engagement in digital-based platforms. By offering users a challenge and a corresponding reward, companies’ services become not only more attractive, but most importantly it can lead users to become frequent customers, thus enhancing a brand’s impact on its customers. However, knowing which coupon to provide can be a rather complex task since each customer profile responds differently to each offer, and frequently offering them bad deals might drag them away from your business. To overcome this problem, machine learning techniques can be used to build data-driven customer profiles and develop better coupon recommendations. | CO1,CO4 | APPLY, analyze |
| 10 | Task5 : REGRESSION: CAR PRICE PREDICTION Imagine a situation where you have an old car and want to sell it. You may of course approach an agent for this and find the market price, but later may have to pay pocket money for his service in selling your car. But what if you can know your car selling price without the intervention of an agent. Or if you are CO 1-4 Apply, Analyze and Evaluate an agent, definitely this will make your work easier. Yes, this system has already learned about previous selling prices over years of various cars. Perform the following tasks on the dataset given: 1. Apply the required preprocessing on the dataset to apply classification / regression. 2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language. 3. Evaluate the model with suitable measures. 4. Perform the comparison analysis of your results. 5. Discuss what your observations and conclusions. | CO1,CO4 | APPLY, analyze |
| 11 | TASK 6: CLASSIFICATION: STUDENT PERFORMANCE Research on the educational field involving machine learning techniques has recently taken a steep growth trajectory. A new term called “Educational Data Mining” has come into existence, i.e., the application of data mining techniques in an educational background aiming to discover hidden trends and patterns about student’s performance. This project aims to develop a prediction mode l for students’ academic performance based on machine learning techniques. The resultant model can be used to identify any student’s performance for a particular subject. | CO1,CO4 | APPLY, analyze |
| 12 | REGRESSION: FLIGHT PRICE PREDICTION The objective of the study is to analyse the flight booking dataset obtained from “Ease My Trip” website and to conduct various statistical hypothesis tests in order to get meaningful information from it. The 'Linear Regression' statistical algorithm would be used to train the dataset and predict a continuous target variable. 'Easemytrip' is an internet platform for booking flight tickets, and hence a platform that potential passengers use to buy tickets. A thorough study of the data will aid in the discovery of valuable insights that will be of enormous value to passengers | CO1,CO4 | APPLY, analyze |
| 13 | TASK 7: CLASSIFICATION: CAMPUS RECRUITMENT Placements hold great importance for students and educational institutions. It helps a student to build a strong foundation for the professional career ahead as well as a good placement record gives a competitive edge to a college/university in the education market. This study focuses on a system that predicts if a student would be placed or not based on the student’s qualifications, historical data, and experience. | CO1,CO4 | APPLY, analyze |
| 14 | Task 7: REGRESSION: BIKE RENTS FOR THE DAY The objective of the study is to analyse the flight booking dataset obtained from CO 1-4 Apply, Analyze and Evaluate “Ease My Trip” website and to conduct various statistical hypothesis tests in order to get meaningful information from it. The 'Linear Regression' statistical algorithm would be used to train the dataset and predict a continuous target variable. 'Easemytrip' is an internet platform for booking flight tickets, and hence a platform that potential passengers use to buy tickets. A thorough study of the data will aid in the discovery of valuable insights that will be of enormous value to passengers. | CO1,CO4 | APPLY, analyze |
| 15 | TASK 8: CLASSIFICATION: CLASSIFICATION: HR ANALYTICS: PREDICT WHO WILL MOVE TO A NEW JOB A company which is active in Big Data and Data Science wants to hire data scientists among people who successfully pass some courses which conduct by the company. Many people signup for their training. Company wants to know which of these candidates are really wants to work for the company after training or looking for a new employment because it helps to reduce the cost and time as well as the quality of training or planning the courses and categorization of candidates. Information related to demographics, education, experience are in hands from candidates signup and enrollment. This dataset designed to understand the factors that lead a person to leave current job for HR researches too. By model(s) that uses the current credentials,demographics,experience data you will predict the probability of a candidate to look for a new job or will work for the company, as well as interpreting affected factors on employee decision. target: 0 – Not looking for job change, 1 – Looking for a job change | CO1,CO4 | APPLY, analyze |
| 16 | Task 8: REGRESSION: MOTOR CYCLE PRICE PREDCITION This dataset contains information about used motorcycles. This data can be used for a lot of purposes such as price prediction to exemplify the use of linear regression in Machine Learning. The columns in the given dataset are as follows: name, selling price, year, seller type, owner, km driven, ex showroom price | CO1,CO4 | APPLY, analyze |
| 17 | TASK 9: CLASSIFICATION: MOBILE PRICE CLASSIFICATION Bob has started his own mobile company. He wants to give tough fight to big companies like Apple,Samsung etc. He does not know how to estimate price of mobiles his company creates. In this competitive mobile phone market you cannot simply assume things. To solve this problem he collects sales data of mobile phones of various companies. Bob wants to find out some relation between features of a mobile phone(eg:- RAM,Internal Memory etc) and its selling price. But he is not so good at Machine Learning. So he needs your help to solve this problem. In this problem you do not have to predict actual price but a price range indicating how high the price is. The last column is the target variable with value of 0(low cost), 1(medium cost), 2(high cost) and 3(very high cost) | CO1,CO4 | APPLY, analyze |
| 18 | Task9 : REGRESSION: AUTO INSURANCE PREDICTION The famous sonar dataset for quick debugging and new application testing purpose Auto Insurance in Sweden. In the following data X = number of claims Y = total payment for all the claims in thousands of Swedish Kronor for geographical zones in Sweden | CO1,CO4 | APPLY, analyze |
| 19 | TASK 10: CLASSIFICATION: DIABETES HEALTH INDICATORS Diabetes is among the most prevalent chronic diseases in the United States, impacting millions of Americans each year and exerting a significant financial burden on the economy. Diabetes is a serious chronic disease in which individuals lose the ability to effectively regulate levels of glucose in the blood, and can lead to reduced quality of life and life expectancy. After different foods are broken down into sugars during digestion, the sugars are then released into the bloodstream. This signals the pancreas to release insulin. Insulin helps enable cells within the body to use those sugars in the bloodstream for energy. Diabetes is generally characterized by either the body not making enough insulin or being unable to use the insulin that is made as effectively as needed. The target variable Diabetes\_binary has 2 classes. 0 is for no diabetes, and 1 is for prediabetes or diabetes. This dataset has 21 feature variables. | CO1,CO4 | APPLY, analyze |
| 20 | TASK 10: REGRESSION: STUDENT GRADE PREDICTION This data approach student achievement in secondary education of two Portuguese schools. The data attributes include student grades, demographic, social and school-related features) and it was collected by using school reports and questionnaires. Two datasets are provided regarding the performance in two distinct subjects: Mathematics (mat) and Portuguese language (por). In [Cortez and Silva, 2008], the two datasets were modeled under binary/five-level classification and regression tasks. Important note: the target attribute G3 has a CO 1-4 Apply, Analyze and Evaluate strong correlation with attributes G2 and G1. This occurs because G3 is the final year grade (issued at the 3rd period), while G1 and G2 correspond to the 1st and 2nd period grades. It is more difficult to predict G3 without G2 and G1, but such prediction is much more useful. | CO1,CO4 | APPLY, analyze |
| 21 | TASK 11: CLASSIFICATION: BANK LOAN DEFAULTER PREDICTION: Predict if a person will be a loan defaulter or not Banks run into losses when a customer doesn't pay their loans on time. Because of this, every year, banks have losses in crores, and this also impacts the country's economic growth to a large extent. In this hackathon, we look at various attributes such as funded amount, location, loan, balance, etc., to predict if a person will be a loan defaulter or not. To solve this problem, MachineHack has created a training dataset of 67,463 rows and 35 columns and a testing dataset of 28,913 rows and 34 columns. | CO1,CO4 | APPLY, analyze |
| 22 | Task11: REGRESSION: BODYFAT PREDICTION Lists estimates of the percentage of body fat determined by underwater weighing and various body circumference measurements for 252 men. This data set can be used to illustrate multiple regression techniques. Accurate measurement of body fat is inconvenient/costly and it is desirable to have easy methods of estimating body fat that are not inconvenient/costly. | CO1,CO4 | APPLY, analyze |