PGP in AI/ML

Regression - Assignment 1

Submission Date: 22:00hrs on 16-11-2021

**Student ID: 2021AIML064**

**Total Marks: 12**

1. The following questions are to enhance your understanding of basic concepts and definitions. You are expected to answer these questions on your own without referring to internet. (5 Marks)
   1. Classify the following use cases into regression or classification algorithms with proper explanations: (0.5\*4 = 2M)
      1. **Facial Recognition:**

It is Classification algorithm. Every face is a combination of distinguishable facial features which can be labelled to identify different persons. Here each face is a discrete dependent variable of combination of facial expressions which are independent variables and can be encoded to construct the face, hence this is a use case for classification algorithm.

* + 1. **Mobile Price Prediction**

Price is dependent variable based on different features like model, makers, configuration etc. Range of price is real numbers hence this is use case for a regression algorithm.

* + 1. **Credit Card Fraud Prediction**

This is Classification algorithm. The dependent variable in this case is a binary output- ‘Yes’ or ‘No’(fraud happened or not) which are discrete values, hence a use case for classification algorithm.

* + 1. **Customer Churn Prediction**

This is Classificationalgorithm. The dependent variable in this case is a binary value whether the customer will stay based on different features.

* 1. Provide differences between regression, classification, and clustering problems. Give an example where each of these models could be used. (3M)

|  |  |  |
| --- | --- | --- |
| **Regression** | **Classification** | **Clustering** |
| Type of supervised learning. | Type of supervised leaning. | Clustering is an unsupervised learning |
| The desired output (labels) as well are given along with the input data and the goal is to predict the relationship between input and output data. | The desired output (labels) as well are given along with the input data and the goal is to predict the relationship between input and output data. | It does not have labelled outputs hence here the goal is to infer the pattern or natural structure present within dataset. |
| The output data also known as dependent variables is a continuous real number values. | The output data also known as dependent variables are discrete values. | The output are discrete values. |
| Example: Forecasting revenue and expenses in an organization.  House price prediction | Example: Based on symptoms, inferring if a person has a COVID 19 or not.  Facial recognition | Example: Market Segmentation  Email Spam Filtering |
| * Linear Regression * Polynomial * Ridge * Lasso * Elastic | * Logistic Regression * K-Nearest Neighbours (kNN) * Decision Trees * Naïve Bayes * Support Vector Machine | • K-Means Clustering  • Agglomerative Hierarchical Clustering (Bottom-Up Approach)  • Divisive Hierarchical Clustering (Top-Down Approach)  • DBSCAN (Density-based Spatial Clustering of Applications with Noise)  • OPTICS  • BIRCH |

1. Solve the following matrix using simultaneous equations in python. (3M)

x + y + z = 6

2y + 5z = -4

2x + 5y -z = 27

1. The following dataset of size n = 31 provides employee-wise salary. The variables are y = salary of an individual employee and x = years of experience of each individual employee .

Find the intercept, coefficient value and error term for this regression problem

in Python. How would you convey (interpret) these results in a business

context?

**Business context:**

1. The salary of a fresher(0 years of experience) is around 26780 units. (Intercept provides the salary at 0 years of experience.)

2. The salary increases with rate of average 9312 units per year. Or, an employee with an additional year of experience will have additional salary of 9312 units. (slope or coef is means 9312 units of salary per year).

3. The model predicts the salary variance against the years of experience pretty well as R-squared stats is close to 99%.

4. The actual salary may have deviation of 3580 units from the predicted salary. (RMSE= 3580 salary units).

**Submissions Instructions:**

Solutions or answers to all questions (excepting programming problems) should be submitted in a word document named ‘Roll\_No.doc’ (Roll\_No is your identity number of this programme).

All python code should be submitted in a jupyter notebook named "Roll\_No.ipynb". Strictly do not submit any rar/zip or any other compressed files.

**NOTE:** For any queries, Email @ [murtuza.dahodwala@wilp.bits-pilani.ac.in](mailto:murtuza.dahodwala@wilp.bits-pilani.ac.in)