**Regression**

**Assignment 1**

**(from 23rd May to 31st May 2024)**

**Q.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. **[6 Marks]**

a. State any four business applications of machine learning. (0.5\*4 = 2M)

**Answer: -** Please find below four business application of machine learning**: -**

1. Medical Diagnosis
2. Self-driving car
3. Hand written digit recognition
4. Email spam detection

b. Explain the difference between supervised learning and unsupervised learning with example. (1 M)

**Answer: -** Please find below difference between supervised learning and unsupervised learning: -

|  |  |
| --- | --- |
| **Supervised learning** | **Unsupervised learning** |
| Input data is labelled | Input data is Unlabelled |
| Solve two types of problem statement   1. Regression, 2- Classification | Solve two types of problem statement   1. Clustering, 2- Dimensionality Reduction |
| Dataset contains one dependent feature & one or multiple independent features | Dataset contains No dependent feature/attribute |
| Build /Fit a model that predict the output as accurately as possible based on independent features | To discover groups of similar examples with in data set. |
| Example: - Salary Prediction, Employability Prediction | Example: - Customer segmentation based on their nature while visiting the store |

c. Provide differences between regression, classification, and clustering problems. Give an example where each of these models could be used.

**Answer:** -

**Regression Problem: -**

1. Regression problem is type of supervised learning problem.
2. In Regression Problem where target/response variables take only continuous values.
3. Types: - Linear Regression, Polynomial Regression & Multilinear Regression
4. Example: - Prediction of used car price, House price prediction

**Classification Problem: -**

1. Classification problem is type of supervised learning problem.
2. In classification problem where target/response variables take discreate values.
3. Example: - Employability Prediction, predicting whether a patient has a particular disease or not.

**Clustering Problem: -**

1. Clustering problem is type of unsupervised learning problem.
2. In clustering problem, we do clustering /grouping or customer segmentations within data set based on multiple independent features (Input).
3. Example: - Customer segmentation based on their nature while visiting the store

(1+1+1 M)

**Q.2**.Solve the following matrix using simultaneous equations in python (2M)

6x+2y−5z=13,3x+3y−2z=13,7x+5y−3z=26

**Answer:** - Program submitted.

**Q.3**.The following dataset of size n = 63 provides total payment (in thousands) for all the claims of Swedish Kronor for geographical zones in Sweden. The variables are y = total payment in thousands and x = number of claims.

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

How would you convey these numbers in a business context?

**Answer:**-   
A fit model calculate model parameters values for above regression problem Intercept:19.994485759114855  
Coefficient:3.4138235600663642  
Error Term (Root Mean Squared Error): 35.94092182307428

* In business context, this intercept value could represent fixed costs or other factors contributing to total payments, independent of the number of claims processed.
* In a business context, this coefficient suggests that for each additional claim processed, total payments are expected to increase by approximately 3.41 units (thousands).
* In a business context, the error term quantifies the accuracy of the regression model's predictions. A lower RMSE indicates that the model's predictions are closer to the actual total payments.

(2+2 = 4M)

(**Submissions:**

Solutions or answers to all questions (excepting programming problems)

should be submitted in a word document named ‘BITS ID\_No.doc’ (BITS \_No is your identity number of this programme).

All python code should be submitted in a Jupyter notebook named

"BITS ID.ipynb".

Strictly, No zip files should be uploaded.

NOTE:

For any queries, mail to *murtuza.dahodwala@wilp.bits-pilani.ac.in*