NAME: KSHITIJ VINOD SALI

CLASS & DIV: TE-A

ROLL NO.: 35059

ML

UA-02

Title: Assignment on Regression Technique.

Problem Statement: Assignment on Regression technique download temperature data from below link:

https://www.kaggle.com/venky73/temperatureofindia?select = temperature.csv.

This data consists of temperature of INDIA averaging the temperature of all places monthly. Temperature value are recorded in CELCIUS.

a) Apply linear regression using suitable library function & predict the monthwise temperature.

b) Assess the performance of regression model using MSE and R-square metrics.

c) Visualize simple regression model.

Course Outcome:

C314448.1: Implement different supervised & unsupervised learning algorithms.

C314448.2: Evaluate performance of machine learning algorithms for real-world applications.

Pre-requisite: Python, Discrete Structure.

Practical Assignment Objective: Understand fundamental elements of machine learning to work on machine learning regression algorithm.

Requirement: Python programming, Jupiter notebook, Google Colab.

Theory:

i) Introduction to regression.

→ Regression is a statistical method used to model the relationship between a dependant variable & one or more independent variables. It helps in predicting the value of dependent variable based on given inputs. In simple linear regression, goal is to fit a straight line that best represents data points, minimizing the error between predicted & actual values.

ii) Types of Universal Regression – Least-Square Method, Model Representation.

→ Universal regression involves predicting a dependend variable using a single independent variable. The least-square method minimizes the sum of squared differences between actual and predicted values, ensuring the best-fit line. Model representation involves defining linear equation $y = mx + b$, where m is the slope & b is the intercept.

iii) Cost Functions: MSE, MAE, R-square.

→ a) MSE: MSE measures the average of squared differences between predicted & actual values. It penalizes larger errors more heavily, making it sensitive to outlines. A lower MSE indicates better model performance.

b) MAE: It calculates the average magnitude of errors between predictions & actual values, without considering their direction. It treats all errors equally, making it robust to outliers. A lower MAE indicates a more accurate model.

c) R-squared: It indicates proportion of variance in dependendt variable explained by independant variable. R^2 values range from 0 to 1, with values closer to 1 indicating a better fit. It helps assess how will the model captures data's variability.

iv) Performance evaluation, optimization of simple linear regression.

→ Performance evaluation in simple linear regression is done using matrics like MSE, MAE & R-squared. Optimization involves adjusting parameters to minimize cost function, which can be done using gradient descent or analytical methods. The goal is to find line that best fit data, predicting out-comes with minimal error.

v) Note on overfitting, underfitting, bias & varience.

→ • Overfitting occurs when a model is too complex & fits the for training data too closely capturing noise & performing poorly on new data.

• Underfitting happens when a model is too complex failing to capture underlying trend.

• Bias refers to errors due to overly simplistic models, while varience refers to sentivity to small functions in training data. The balance between

bias & variance is crucial for generalization.

Algorithm: Linear regression is used to predict a quantitative response Y from predictor variable X.
Mathematically, we write as:

$$y = a + bx.$$

where,

$$b = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$

$$a = \frac{n\Sigma y - b(\Sigma x)}{n}$$

Here, x & y are two variables on regression line.
b = Slope of the line.
a = y-intercept of the line.
x = Independent variable from dataset.
y = Dependent variable from dataset.

Input: We have given temperature data as an input to regression model to calculate different outputs.

Output: Students are able to design regression models & predict the different temperature values.

Inference: Thus, we have studied about regression and assess the performance of regression models using MSE, MAE and R-square matrics.