ICA2 – Regression Analysis

Deadline: September 12, by 1:50 PM

Objective: The primary objective of this analysis is to learn about **regression analysis** by applying two different regression models, **Linear Regression** and **Decision Tree Regression**, to predict the actual mean temperature. Through this process, we aim to understand how different models handle regression tasks and assess their predictive performance using evaluation metrics like Mean Squared Error (MSE) and R-squared.

Dataset Description: This dataset contains daily weather data, including actual and historical temperatures, precipitation levels, and record values.

Q1) Build and train a **linear regression model** using the provided dataset in the "KCLT.csv" file, which contains temperature data. Focus on predicting the values in the actual_mean_temp column using actual_max_temp as the feature.

- Set actual_max_temp as the independent variable (feature) and actual_mean_temp as the dependent variable (target).
- Split the dataset into training and testing sets, using an 80-20 ratio. Use 42 for the random state.
- Use the trained model to predict values for the test set.
- Compute and display the Mean Squared Error (MSE) and R-squared (R^2) to assess the model's performance.
- Output the model's coefficients and intercept.
- Create a DataFrame that compares the actual test set values with the predicted values, displaying the first few rows.
- Generate a scatter plot of the actual actual_max_temp values against actual_mean_temp, overlaying the regression line that represents the predicted values.
- Finally, analyze the model's performance based on the obtained results.

Q2) build and evaluate a **Decision Tree Regressor** model to predict the average temperature (actual_mean_temp) based on various temperature-related features. You will use the provided dataset to train and test your model and assess its performance..

1. Use the following columns as features:

- actual_min_temp
- actual_max_temp
- average_min_temp
- average_max_temp
- record_min_temp
- record_max_temp
- record_min_temp_year
- record_max_temp_year

- actual_precipitation
- average_precipitation
- record_precipitation

And use actual_mean_temp as the dependent variable (target).

- Split the dataset into training (70%) and testing (30%) sets. Use 42 for the random state.

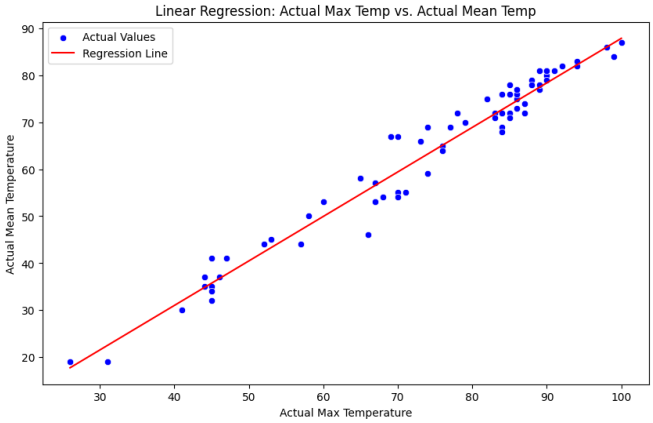
2. Train a Decision Tree model on the training data.
3. Predict on the test set, then calculate and display the Mean Squared Error (MSE) and R-squared (R^2) values.
4. Create a DataFrame to compare actual versus predicted values and print it.
5. Finally, analyze the model's performance based on the obtained results.

Expected OUTPUT:

Q1)

```
Mean Squared Error: 11.302873810178504
R-squared: 0.9608178883866981
Coefficients: [0.94674867]
Intercept: -6.854358750464577
     Actual  Predicted
193      30  31.962337
33       77  74.566027
15       76  73.619278
309      69  72.672530
57       72  75.512776
..      ...        ...
203      53  56.577802
82       71  73.619278
94       70  67.938786
192      34  35.749332

325      64  65.098540
```



Q2) `Mean Squared Error: 3.9454545454545453`

```
R-squared: 0.9862275523383303
     Actual  Predicted
193      30       31.0
33       77       78.0
15       76       76.0
309      69       70.0
57       72       72.0
..      ...        ...
158      47       47.0
229      25       29.0
360      85       84.0
209      45       47.0
328      73       71.0


[110 rows x 2 columns]
```

Instructions to save your Colab notebook and submit it via Blackboard

1. Run and Save Your Notebook:
• Click on the notebook name at the top and rename it to ICA2-YourGlobalID.ipynb.
• Ensure that all cells in your notebook have been executed so that the expected outputs are visible.
• Go to File > Save to save your progress.
2. Download the Notebook:
• Go to File > Download > .ipynb to save the notebook to your computer.
3. Submit via Blackboard:
• Log in to Blackboard, find the assignment link, upload the ICA2-YourGlobalID.ipynb file, and click Submit.