Title: Weather Forecasting Using Multivariate Linear Regression

---

Objective

To design and implement a multivariate linear regression model that predicts future weather parameters such as temperature or humidity based on historical meteorological data, including variables like air pressure, humidity, wind speed, and previous temperature records.

Theory

Multivariate linear regression is a supervised machine learning technique that models the linear relationship between a dependent variable and multiple independent variables. In the context of weather forecasting, it can be used to predict continuous weather outcomes based on several climatic features. Unlike uni-variate regression, which uses only one feature, multivariate regression considers multiple variables to improve prediction accuracy.

This approach assumes that weather patterns follow linear trends with respect to time and related atmospheric variables. Though simplistic compared to complex models like neural networks or ensemble methods, linear regression offers an interpretable and computationally efficient forecasting method.

Key Concepts

- **Multivariate Linear Regression**: A model that predicts a continuous target variable using multiple input features. The general form is:

  $$Y = \beta0 + \beta1X1 + \beta2X2 + \cdots + \beta nXn + \epsilon$$

  where Y is the dependent variable and X1,X2,...,Xn are the features.
- **Mean Squared Error (MSE)**: Measures the average squared difference between predicted and actual values.
- **Root Mean Squared Error (RMSE)**: The square root of MSE, which represents the standard deviation of prediction errors.
- **R^2 Score (Coefficient of Determination)**: Indicates how well the model explains the variance in the target variable.

Procedure

1. Data Collection

- Obtain a historical weather dataset from sources like Kaggle, NOAA, or other meteorological databases.
- Example features: Temperature (current and lagged), Humidity, Wind Speed, Atmospheric Pressure, Cloud Cover, etc.
- Target variable: Next day's temperature or humidity.

2. Data Preprocessing

- Load data using Python (Pandas).
- Handle missing values (e.g., using interpolation or deletion).
- Encode categorical variables (if any) using one-hot encoding.
- Normalize/scale the features if needed to ensure uniform contribution to the model.

3. Exploratory Data Analysis (EDA)

- Plot pairwise relationships between variables using scatterplots and heatmaps.
- Use correlation matrices to understand relationships between input variables and the target.

4. Model Building

- Split the dataset into training and testing sets (e.g., 80:20 ratio).
- Implement multivariate linear regression using `LinearRegression` from `sklearn.linear_model`.
- Train the model on the training data.

5. Model Evaluation

- Make predictions on the test data.
- Calculate metrics: MSE, RMSE, and R^2 score.
- Analyze residuals (difference between actual and predicted values).

6. Visualization

- Plot actual vs. predicted values to visually assess model performance.
- Optionally visualize coefficients to interpret feature importance.

Conclusion

In this experiment, we successfully developed and evaluated a multivariate linear regression model for weather forecasting. The model predicted future weather conditions, such as temperature, based on historical atmospheric data. The performance was assessed using R^2 and RMSE, demonstrating the model's potential for practical forecasting. Although linear models have limitations in capturing complex non-linear relationships, this method offers a transparent and fast baseline approach for weather prediction tasks.