Week-06-L-02

Agricultural Statistics in Practice

Multivariate Analysis

Multivariate Regression Analysis

Dr. Amandeep Singh

Imagineering Laboratory
Indian Institute of Technology Kanpur

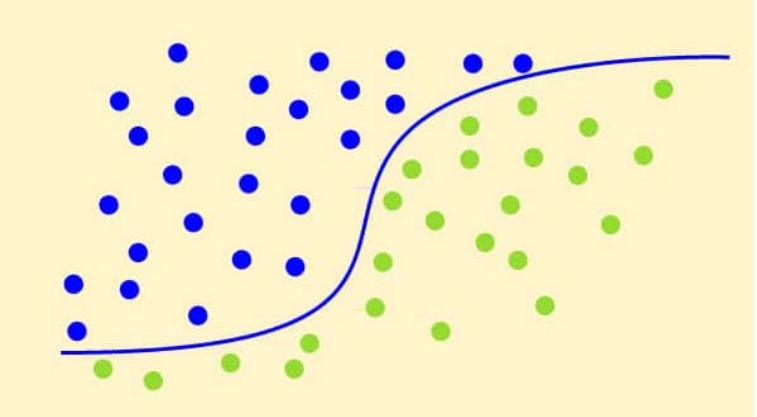
What is multivariate regression?

(2 or more dep. var.)

Multiple Regression

(1- dep. var.)

- Multivariate regression is used when statistically analyzed data involves multiple response variables and explanatory variables.
- It measures the linear relationship between these multidimensional variables through correlation and predicts the behavior of the response variable based on predictor variables.
- Multivariate regression is commonly employed as a supervised algorithm in machine learning for predicting the behavior of dependent variables using multiple independent variables.



Source: voxco.com

Characteristics of multivariate regression

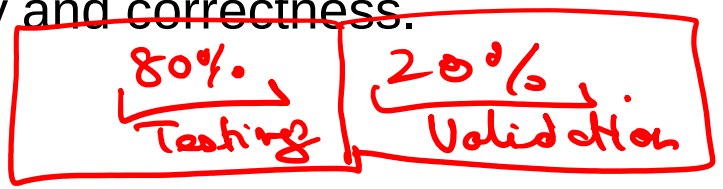
1. Multivariate regression provides a comprehensive perspective on the relationship between multiple variables from different angles.
2. It enables the prediction of response variable behavior based on the movements of predictor variables.
3. Multivariate regression finds applications in diverse fields including
   - machine learning,
   - economics,
   - scientific research,
   - medical studies, etc.

Example

- An agricultural expert investigates the factors contributing to crop damage in a specific region. Data is collected on variables such as recent climate changes, water supply, irrigation methods, pesticide usage, etc., to understand the causes of issues like blackening, lack of fruit yield, and premature drying of crops.
- In this scenario, the expert recognizes the need to analyze multiple independent variables that influence the dependent variables, which represent the crop conditions. Single regression analysis would be inadequate, making multivariate regression a suitable approach to uncover the underlying relationships.

Steps to achieve multivariate regression:

- **Feature Selection**: Identify the key feature that strongly influences the dependent variable.
- **Feature Normalization**: Scale the selected features within a specific range, such as 0-1, for easier analysis.
- **Hypothesis Formulation**: Define the hypothesis as the predicted value of the response variable.
- **Loss Function Selection**: Choose a loss function to measure the deviation of the hypothesis from the actual values.
- **Minimization of Cost and Loss**: Use optimization algorithms, like gradient descent, to minimize the cost and loss functions.
- **Hypothesis Testing**: Validate the formulated hypothesis by testing it against a separate test dataset for accuracy and correctness.



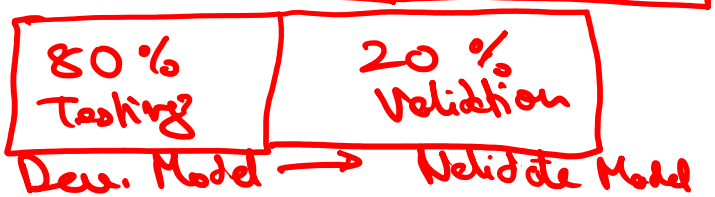
Features and Challenges

Features:

- Multivariate regression allows for the identification of relationships between multiple variables or features.
- It helps determine the correlation between independent and dependent variables.

Challenges:

- Multivariate regression involves complex mathematical calculations.
- The model output can be challenging to analyze.
- The presence of errors in the output can impact the accuracy of the model.
- Multivariate regression performs better with larger datasets compared to smaller ones.



Thank You