

 image-20251002-045406.png

Part 2

Machine Learning Linear Regression Project for Beginners in Python to Build a Multiple Linear Regression Model on Soccer Player Dataset.

What will you learn?

- What is Multiple Linear Regression?
- General Linear Regression Model
- Matrix Representation for General Linear Regression Model
- Matrix Representation of Least Squares
- Understanding types of predictive variables
- F-test
- Coefficient of multiple determination
- Adjusted R- squared
- What are scatterplots?
- What is a correlation matrix?
- Understanding multicollinearity.

- Anova Partitioning
- Diagnostic and Remedial Measures
- What are Indicator Variables?
- Various criteria for model selection such as R^2 , mallows cp Criterion, AIC/SBC Criterion, Press Criterion
- Building a Multiple Linear Regression model

Project Description

Overview

We have started our journey of understanding the background and basics of regression in the first project of this series. We understood the fundamentals of regression and created a simple linear regression model.

In this project, we will get familiarised with multiple linear regression. Unlike linear regression, multiple linear regression is used to estimate the relationship between two or more independent variables and one target / dependent variable.

Before starting this project, please do visit the first project of the series [Linear Regression Model Project in Python for Beginners Part 1](#).

Aim

To build a multiple linear regression model in python.

Data Description

The dataset used is the soccer player dataset. It has information about various players from different clubs, and it provides data over ten features with a number of goals as the target variable.

Tech Stack

- Language: Python
- Libraries: numpy, pandas, statsmodel, seaborn, matplotlib, sklearn, scipy

Approach

1. Import the required libraries and dataset

2. Check for the correlation between features
3. Plot a graph for correlations
4. Remove the weakly correlated and highly multicollinear variables
5. Perform train test split on the dataset
6. Fit the multiple linear regression model
7. Convert categorical variables into dummy/indicator variables
8. Plot the results