Fundamentals of simple linear regression in python for beginners
Project Overview

Overview

Regression is one of the foundational techniques in Machine Learning. Being one of the most well-understood algorithms, beginners always struggle to understand some fundamental terminology related to regression. In this series of projects, we try to give you basic ideas of underlying concepts with the help of practical examples. If you are starting your career or want to brush up on your knowledge of regression, this course is made up for you. This project begins by introducing some simple real-life examples for regression. From a brief introduction to most of the concepts used in regression to hands-on experience, this project will give you enough understanding to apply those in real-world problems. With the help of the background developed, you will code your regression model in python.

Aim

To give a gentle introduction to the fundamentals of regression and build a simple 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: pandas, statsmodel, seaborn, matplotlib, sklearn, scipy

Approach

This project starts with a real-life example for regression analysis, with an introduction to simple and multiple linear regression. Building the statistical foundation for the regression, it gives you a brief idea of the formula of regression. With this background, the first regression model in python is built. Going through the interpolation and extrapolation also explains errors in regression and Lurking variables. The point estimators of mean and variance and distributions of underlying parameters are also discussed. The coefficient of determination is also known, and R squared is briefly

explained. The project ends with diagnostics and remedial measures for regression with a practical explanation.

Project Takeaways

1. What is Regression?
2. Types of Regression
3. What is Mean, Variance, and Standard Deviation?
4. Correlation and Causation
5. What are Observational and Experimental data?
6. Formula for Regression
7. Building a Simple Linear Regression model
8. Understanding Interpolation and Extrapolation
9. What are Lurking Variables?
10. Derivation for Least Square Estimates
11. The Gauss Markov Theorem
12. Point estimators of Regression
13. Sampling distributions of Regression coefficients
14. F- Statistics
15. Anova Partitioning
16. Coefficient of Determination(R-Squared)
17. Diagnostic and Remedial Measures