Linear Regression

What is Linear Regression?

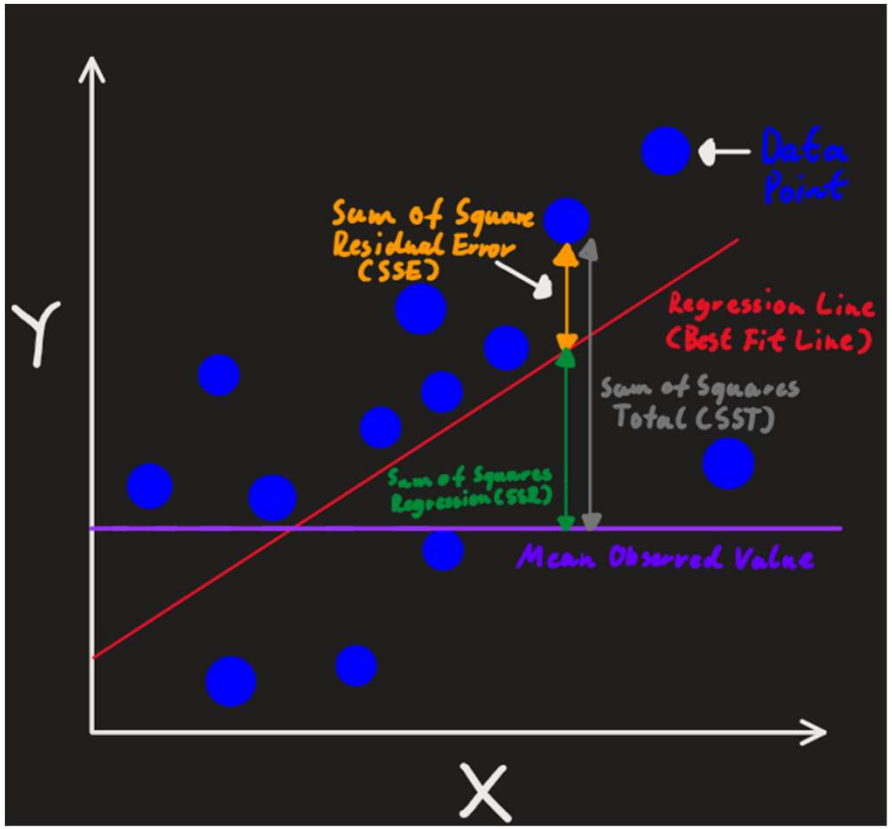
- Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. The goal is to find a linear equation that best fits the data.

Purpose of Linear Regression

- The main purpose is to predict outcomes based on quantitative data and to determine the strength of predictors.

Application Example

- In wine quality assessment, linear regression can predict the quality of wine based on features like alcohol content.



The graphic illustrates linear regression, highlighting the concepts of SSE (Sum of Squares Error), SST (Sum of Squares Total), SSR (Sum of Squares Regression), the regression line, and the mean of the response variable.

Implementation

1. Linear Regression Model

- Extracting alcohol content as a feature and quality as the target.
- Creating, fitting and predict the Linear Regression model using scikit-learn.

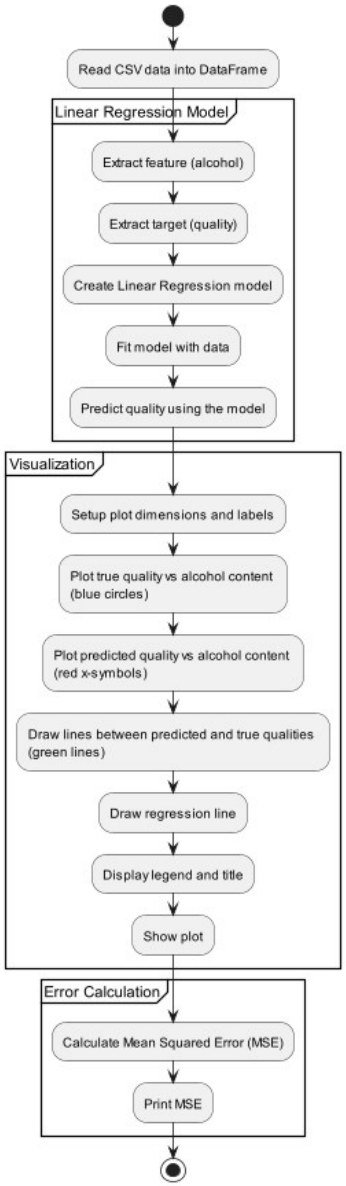
3. Visualization:

- Plotting real vs predicted quality to visualize the model's performance.
- Highlighting residuals to understand prediction errors.

3. Error Calculation:

- Computing the Mean Squared Error (MSE) to quantify model accuracy.

Linear Regression - algorithm description



Results and Analysis

Visualization Insights

- Visualization showcases the model's predictions (red x-symbols) versus actual qualities (blue circles) of wine.
- Residuals, shown by green lines, highlight the prediction errors.
- Regression Line, the red line indicates the general trend, correlating wine quality with alcohol content.

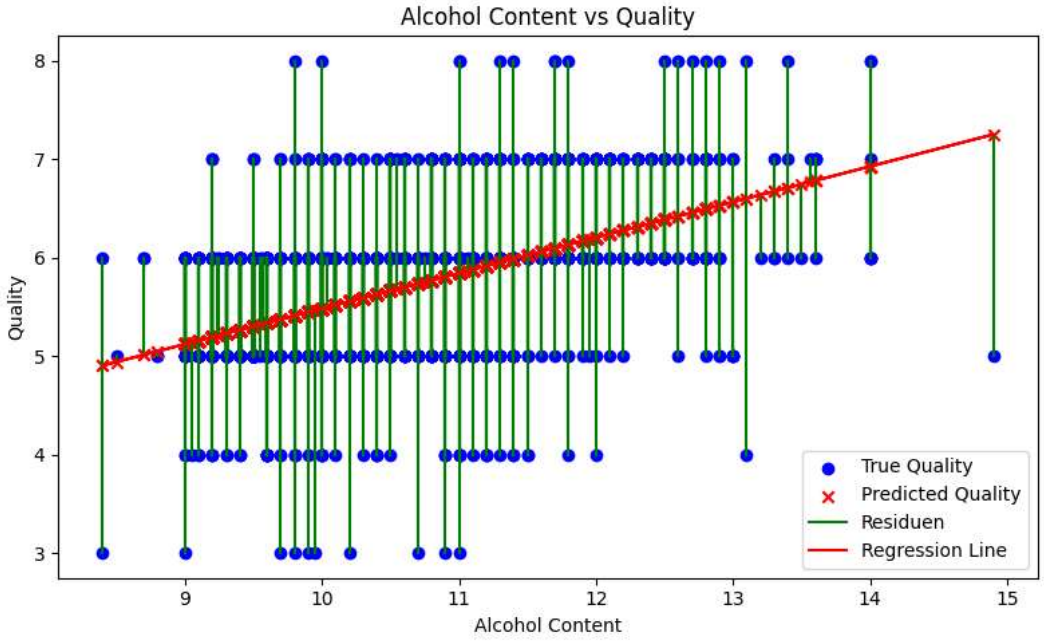
Mean Squared Error (MSE)

- MSE measures the average squared difference between observed and predicted values.
- Calculation: $MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2$

Interpretation:

- An MSE of 0.504 indicates a moderate fit, suggesting the model generally captures the trend but could be improved for greater accuracy. The plot visually reinforces this, showing areas where the prediction errors are larger, which indicates that improvements could be made to achieve a more accurate and reliable model.

Output:




```
Mean Squared Error (MSE): 0.503984025671457
```