Linear Regression

What is linear regression?

Linear regression analysis is **used to predict the value of a variable based on the value of another variable**. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

There are two main types of linear regression:

- **Simple Linear Regression**
- **Multiple Linear Regression**

Simple Linear Regression:

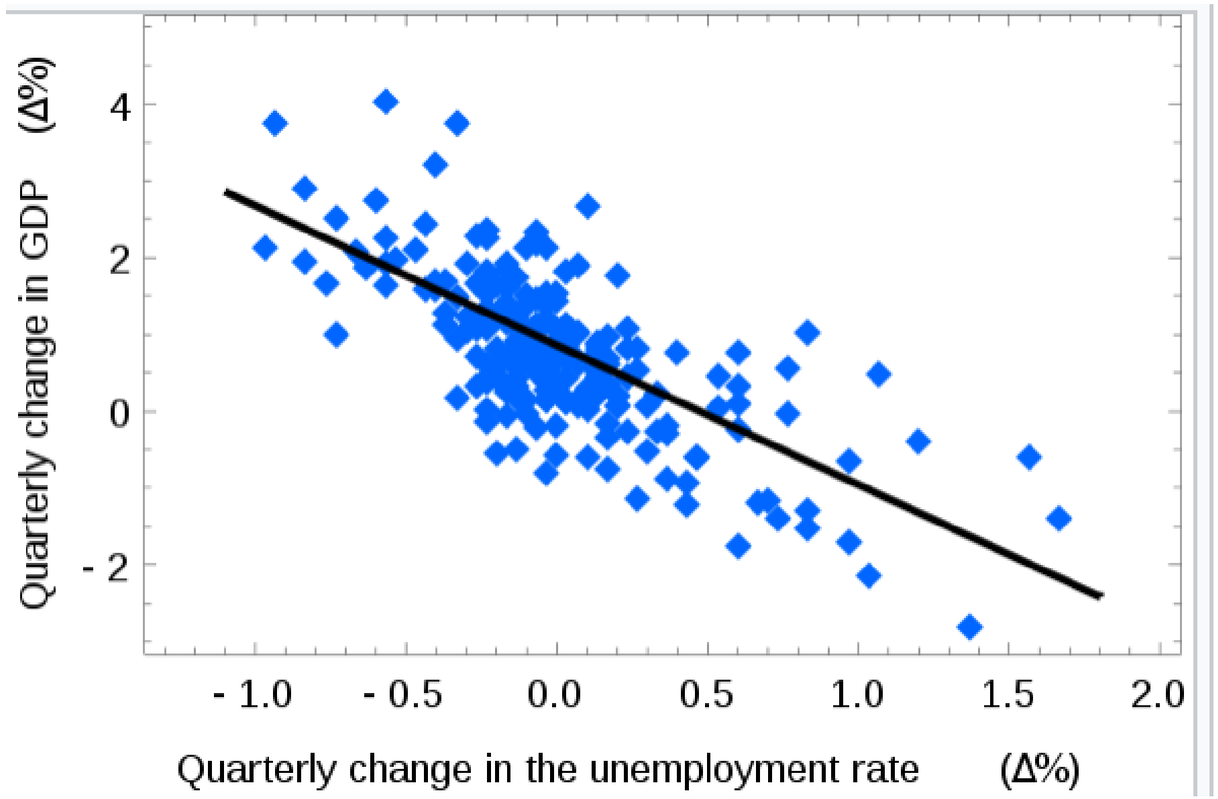
Simple Linear Regression is a type of regression algorithm that models the relationship between a dependent variable and a single independent variable

$$y = a_0 + a_1 x + \epsilon$$

Loss function:

$$\text{MSE} = \frac{1}{N}\sum_{i}^{N}(Y_i - \hat{Y}_i)^2$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|Y_i - \hat{Y}_i|$$



Multiple Linear Regression:

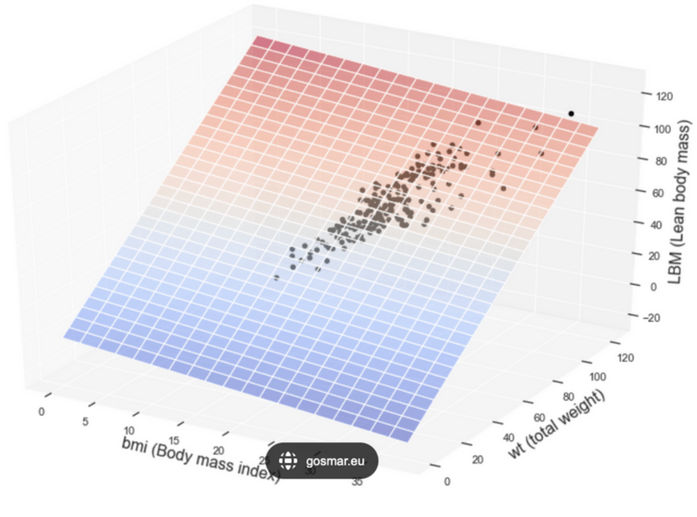
This involves more than one independent variable and one dependent variable. The equation for multiple linear regression is

$$y = \beta_0 + \beta_1 X + \beta_2 X + ...\beta_n X$$

Loss function:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$



MSE is commonly used because it is easier to optimize mathematically, while MAE is more robust to outliers

R-Squared (R2):

- **Definition**: R-squared measures the **proportion of variance** in the dependent variable (*target*) that can be **explained by the independent variables** (*predictors*) in the model.
- **Range**: It varies from **0 to 1**, where 0 indicates that the model does not explain any variability, and 1 indicates that it explains all the variability.
- **Interpretation**: Higher R-squared values suggest a better fit, but it doesn't necessarily mean the model is a good predictor in an absolute sense.
- **Pros**: Easy to understand and widely used.
- **Cons**: It doesn't account for the number of predictors or overfitting.

Adjusted R-Squared:

- **Definition**: Adjusted R-squared addresses a limitation of R-squared, especially in **multiple regression models** (models with more than one independent variable).
- **Adjustment**: While R-squared tends to increase as more variables are added to the model (even if they don't significantly improve the model), **Adjusted R-squared penalizes the addition of unnecessary variables**.
- **Formula**: Adjusted R-squared is calculated as:

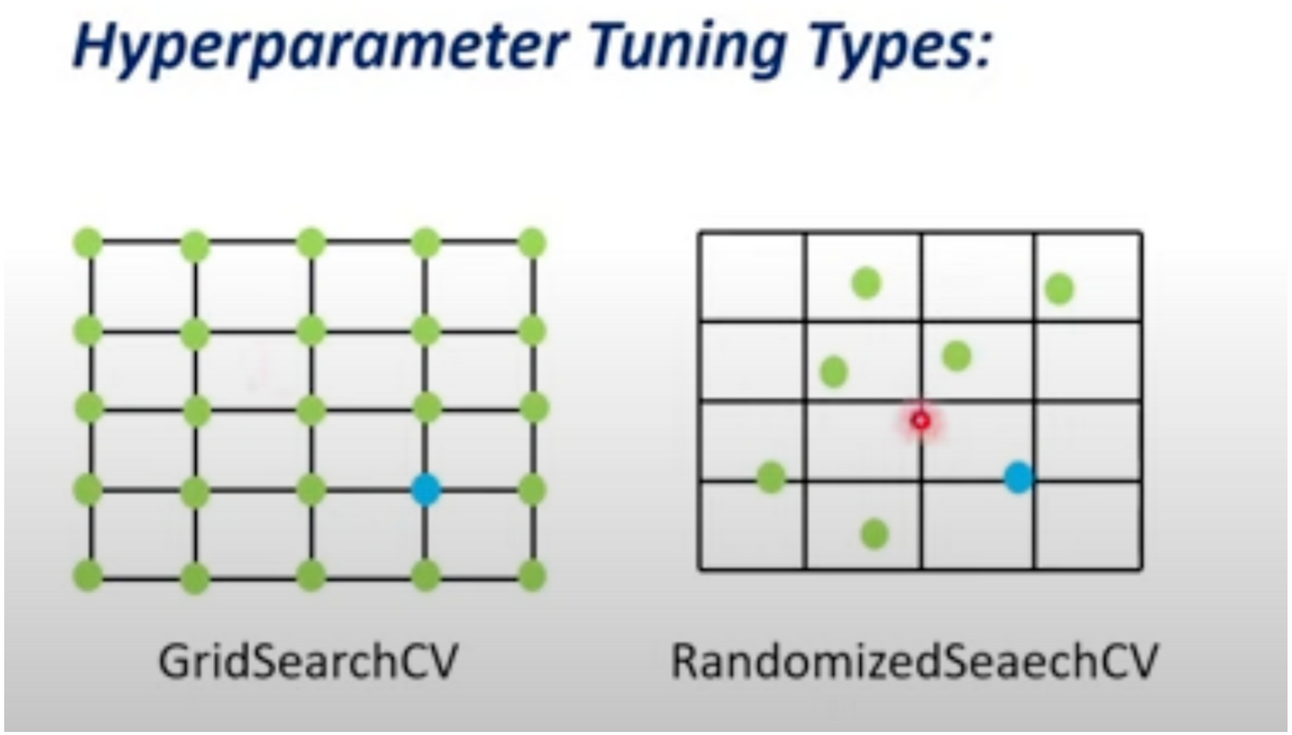
$$\text{Adjusted R-squared} = 1 - \left(\frac{(1 - R^2)(n - 1)}{n - k - 1}\right)$$

Adjusted R-squared is a better choice when comparing models with different numbers of independent variables

Hyperparameter Tuning:

- GridSearchCV
- RandomizedSearchCV

Hyperparameters are parameters whose values control the learning process. These are adjustable parameters used to obtain an optimal model. External Parameters.



Thank You.!!