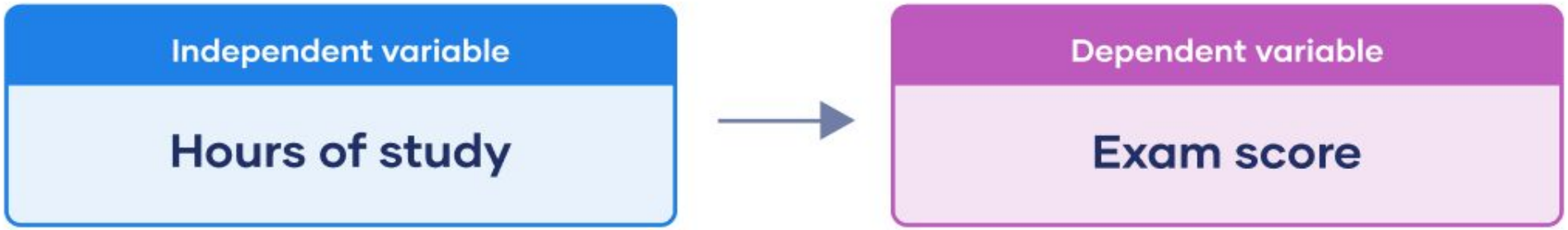
Agenda

- Linear Regression
- Evaluation Metric for Linear Regression
- Standardization
- Training a Regressor
- Different Regression Models
- Comparison of results

Linear Regression

Linear regression is a **supervised** machine learning algorithm used for predicting a **continuous output** (**dependent variable**) based on one or more **input features** (**independent variables**).

Dependent vs independent Variable

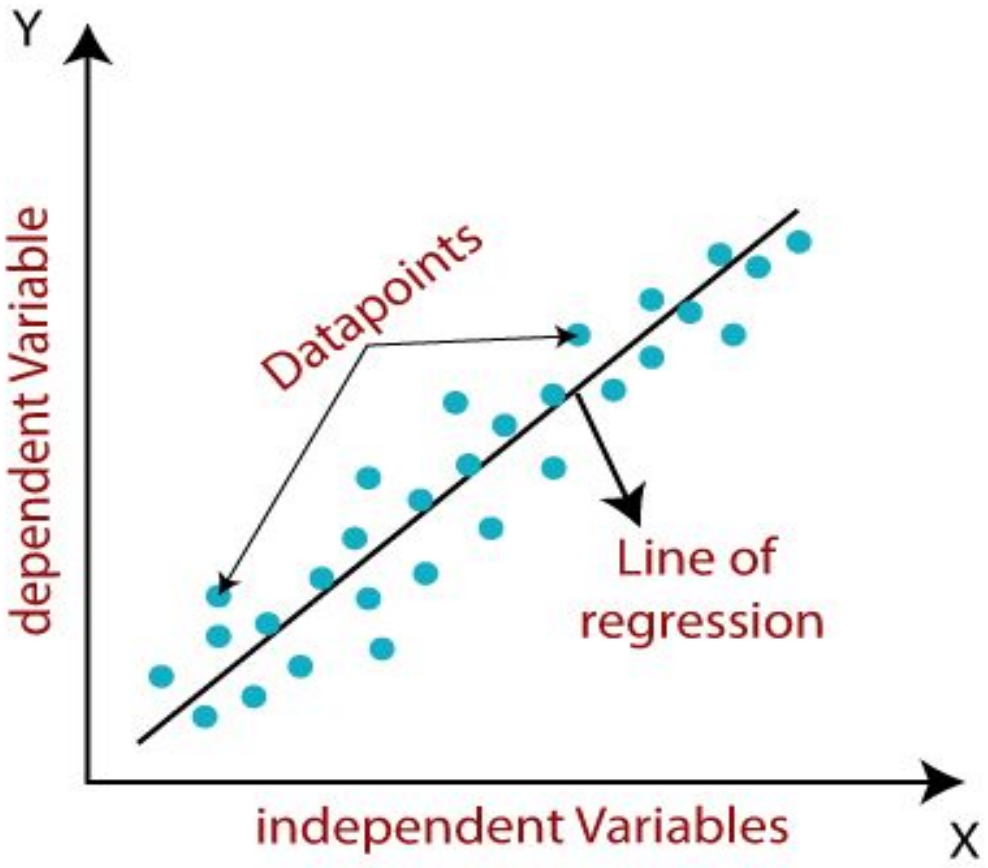


Why use Linear Regression?

- Widely used in various fields, including **finance**, **economics,** and **data science**.
- Great for making **predictions** and understanding relationships **between variables**.

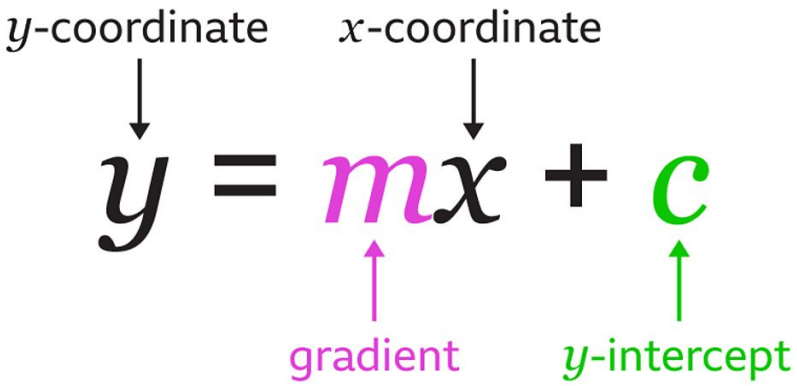
How does Linear Regression work?

It finds the best-fitting line that represents the relationship between X and Y.



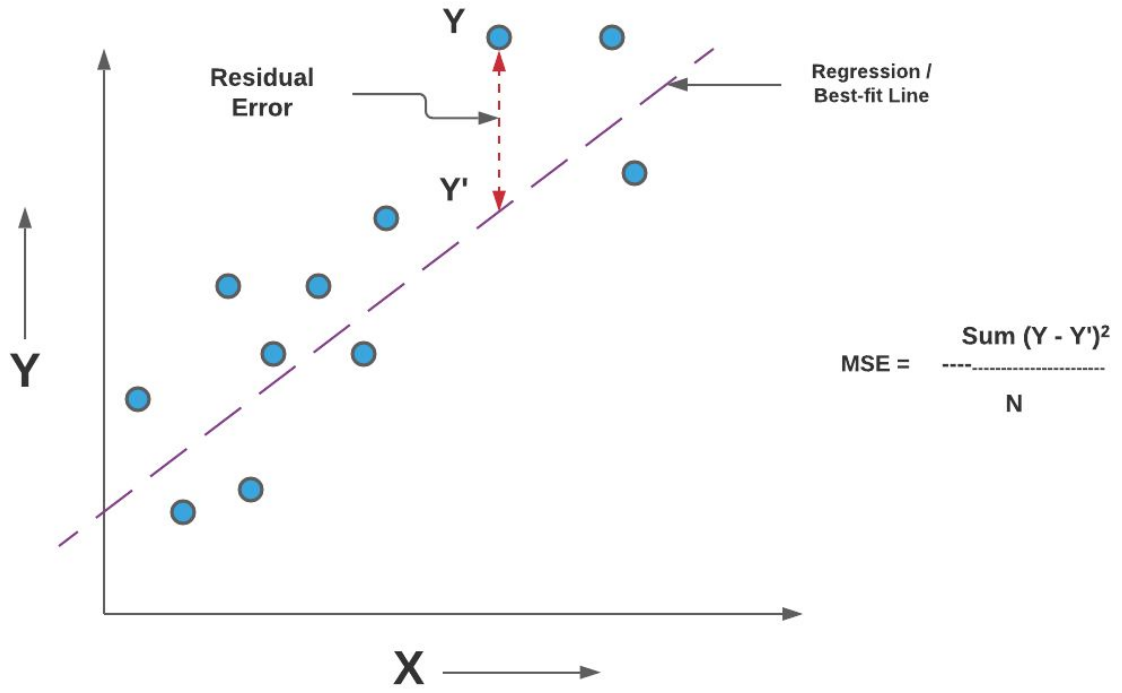
Understanding the Linear Equation

The linear equation: **y = mx + c**
y: The dependent variable (output/prediction we want to make)
x: The independent variable (input/feature)
m: The slope of the line (how much y changes when x changes)
b: The y-intercept (the value of y when x is 0)



Mean Squared Error (MSE)

The **Mean Squared Error (MSE)** is a common evaluation metric used in regression tasks.
It measures the average squared difference between the actual and predicted values.



Standardization

Standardization is a type of data preprocessing that makes different features have a **similar scale.**

Example

Suppose we have data on students' height in **centimeters** and weight in **kilograms.**

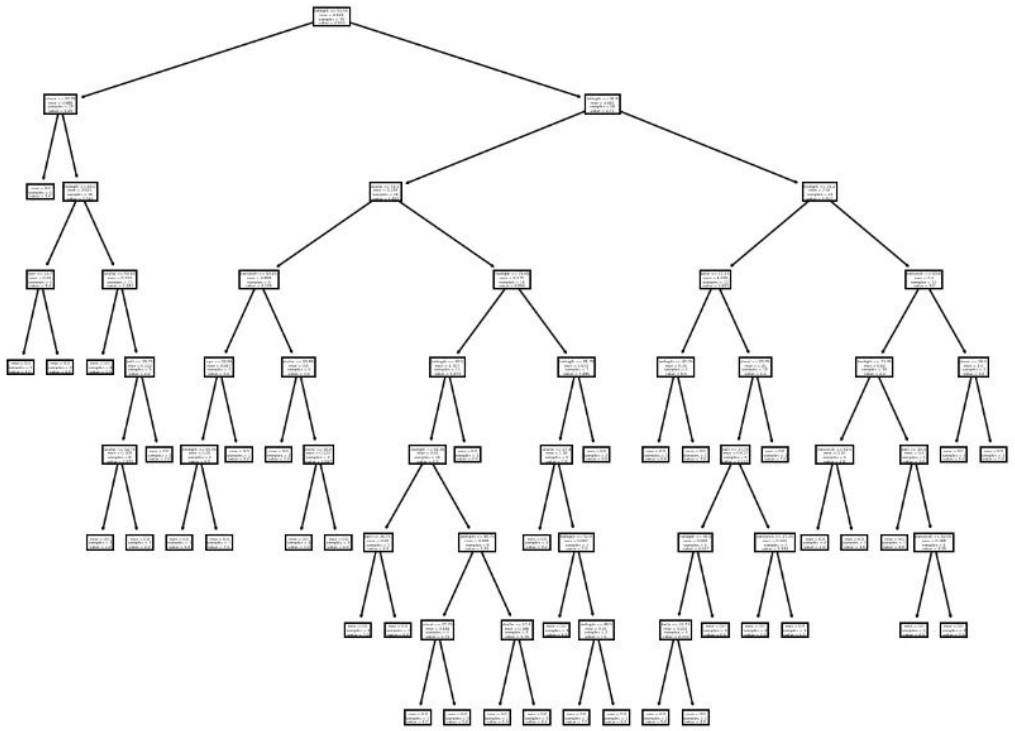
Heights might range from **150 cm to 180 cm,** while weights might range from **40 kg to 80 kg**.

Standardization scales both height and weight so they have similar ranges, like between **0** and **1**.

Different Regression Techniques

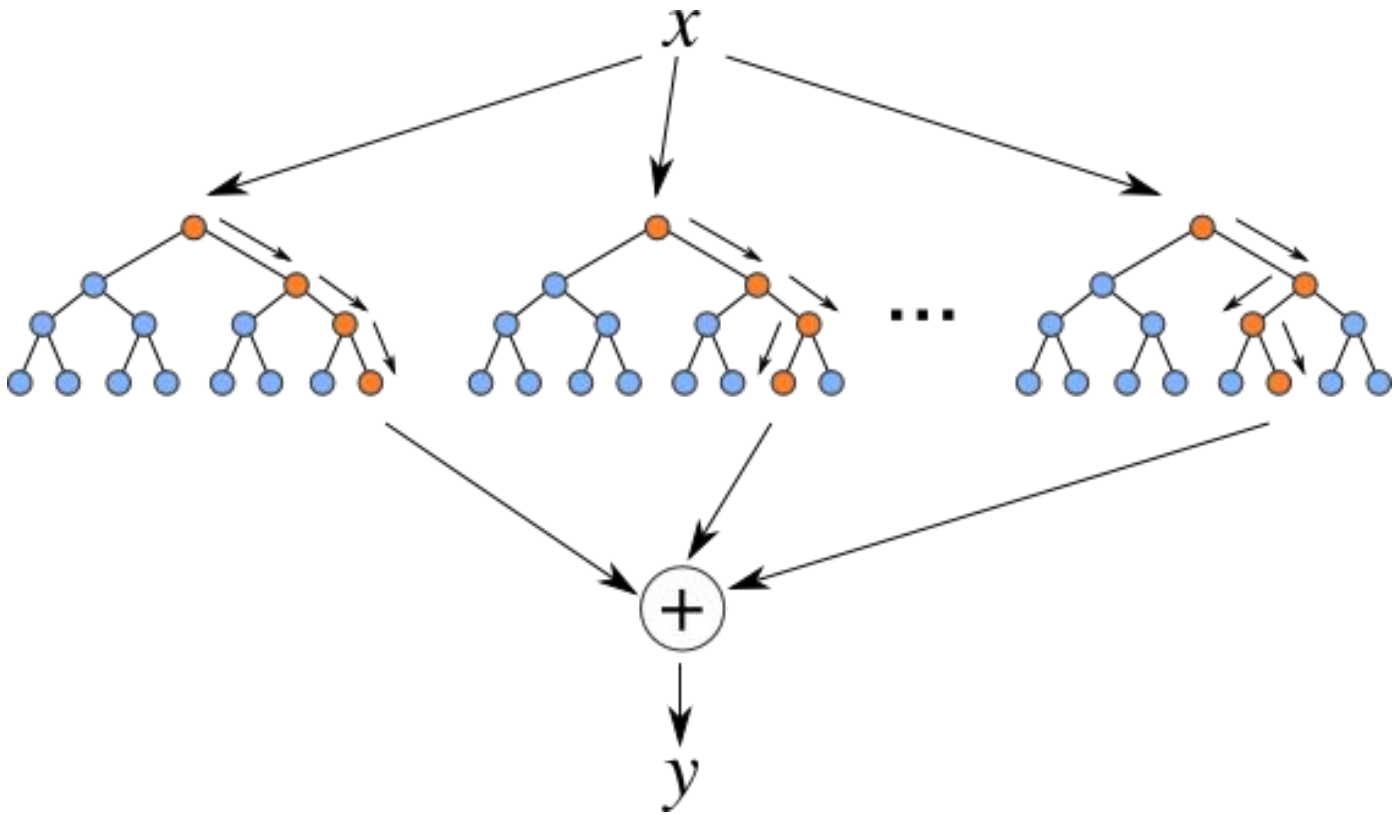
Decision Tree Regressor

Decision tree regression involves **partitioning the data into subsets** based on the values of **independent variables** and predicting the average of the target variable for each subset.



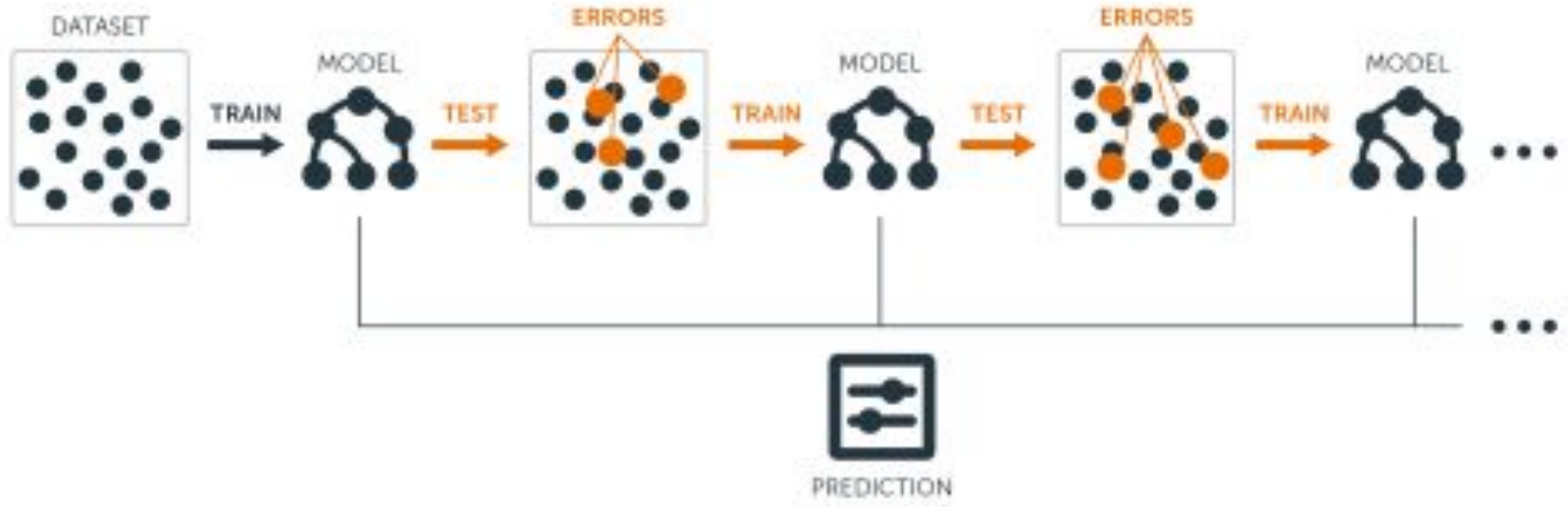
Random Forest Regressor:

Random Forest is learning method that creates **multiple decision trees** during training and outputs the **average prediction** (for regression) from all individual trees.



Gradient Boosting Regressor

Gradient Boosting is learning technique that builds multiple decision trees sequentially, each one correcting the errors of its predecessor.



Comparison of Different regressors

- **Linear Regression:** For simple and linear relationships.
- **Decision Tree:** When dealing with non-linear patterns and interpretability is a priority.
- **Random Forest:** For improved performance and robustness against outliers.
- **Gradient Boosting:** When high accuracy is crucial, and longer training times are acceptable.