



# Welcome to the **Co**Grammar Machine Learning Introduction

The session will start shortly...

Questions? Drop them in the chat. We'll have dedicated moderators answering questions.



## Data Science Session Housekeeping

---

- The use of disrespectful language is prohibited in the questions, this is a supportive, learning environment for all - please engage accordingly.  
**(Fundamental British Values: Mutual Respect and Tolerance)**
- No question is daft or silly - **ask them!**
- There are **Q&A sessions** midway and at the end of the session, should you wish to ask any follow-up questions. Moderators are going to be answering questions as the session progresses as well.
- If you have any questions outside of this lecture, or that are not answered during this lecture, please do submit these for upcoming Academic Sessions. You can submit these questions here: [Questions](#)

## Data Science Session Housekeeping cont.

---

- For all **non-academic questions**, please submit a query:  
[www.hyperiondev.com/support](http://www.hyperiondev.com/support)
- Report a **safeguarding** incident:  
[www.hyperiondev.com/safeguardreporting](http://www.hyperiondev.com/safeguardreporting)
- We would love your **feedback** on lectures: [Feedback on Lectures](#)

# Skills Bootcamp

## 8-Week Progression Overview

### Fulfil 4 Criteria to Graduation

#### ✓ Criterion 1: Initial Requirements

Timeframe: First 2 Weeks

Guided Learning Hours (GLH):

Minimum of 15 hours

Task Completion: First four tasks

**Due Date: 24 March 2024**

#### ✓ Criterion 2: Mid-Course Progress

**60** Guided Learning Hours

Data Science - **13 tasks**

Software Engineering - **13 tasks**

Web Development - **13 tasks**

**Due Date: 28 April 2024**

# Skills Bootcamp Progression Overview

## ✓ Criterion 3: Course Progress

Completion: All mandatory tasks,  
including Build Your Brand and  
resubmissions by study period end  
Interview Invitation: Within 4 weeks  
post-course  
Guided Learning Hours: Minimum of  
112 hours by support end date  
(10.5 hours average, each week)

## ✓ Criterion 4: Demonstrating Employability

Final Job or Apprenticeship  
Outcome: Document within 12  
weeks post-graduation  
Relevance: Progression to  
employment or related  
opportunity

# CoGrammar

## Machine Learning Intro

April 2024



# Learning objectives

## **Understand and implement simple linear regression models using Python and scikit-learn.**

- ❖ Define simple linear regression and its purpose
- ❖ Interpret the mathematical equation and assumptions of simple linear regression
- ❖ Implement and evaluate simple linear regression models using Python



# Introduction to Machine Learning

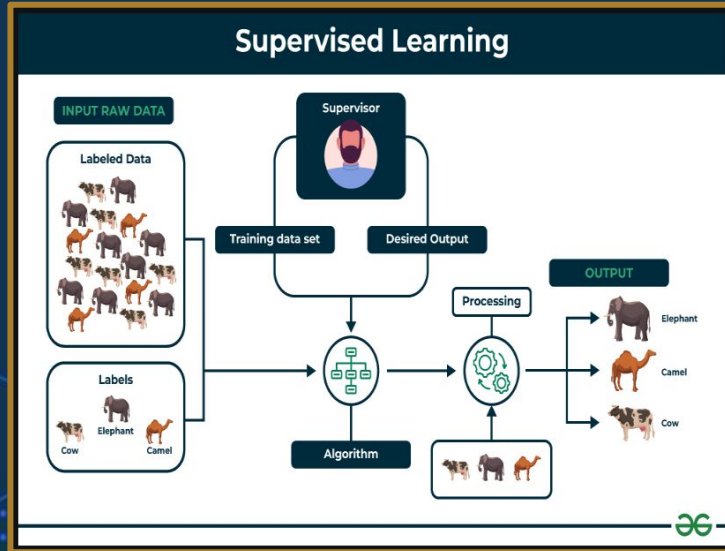


- ❖ Machine learning is a way of teaching computers to learn and improve from experience without being explicitly programmed.
- ❖ It allows computers to automatically learn and adapt based on data.



# Types of machine learning

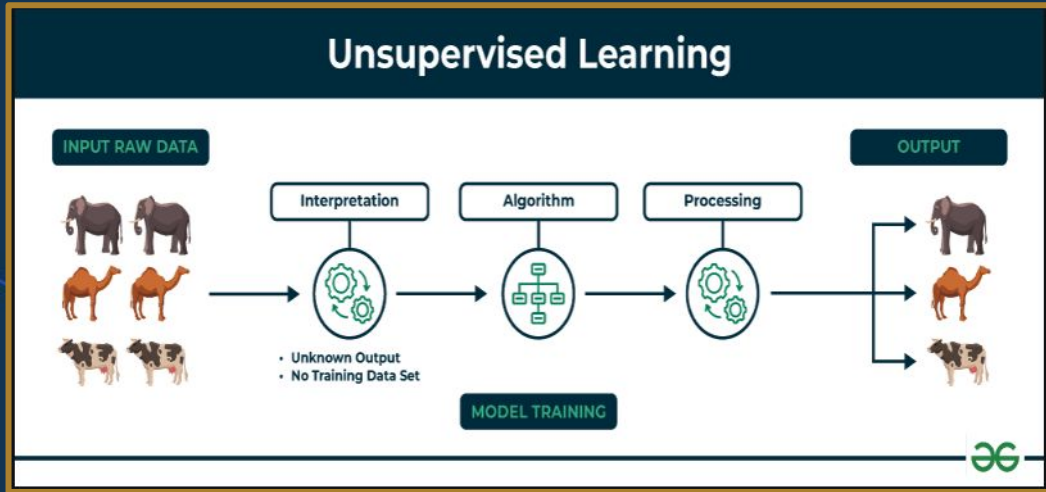
- ❖ **Supervised learning:** The computer learns from labelled data, where both input and output data are provided.



Source: [geeksforgeeks](https://www.geeksforgeeks.org/)

# Types of machine learning

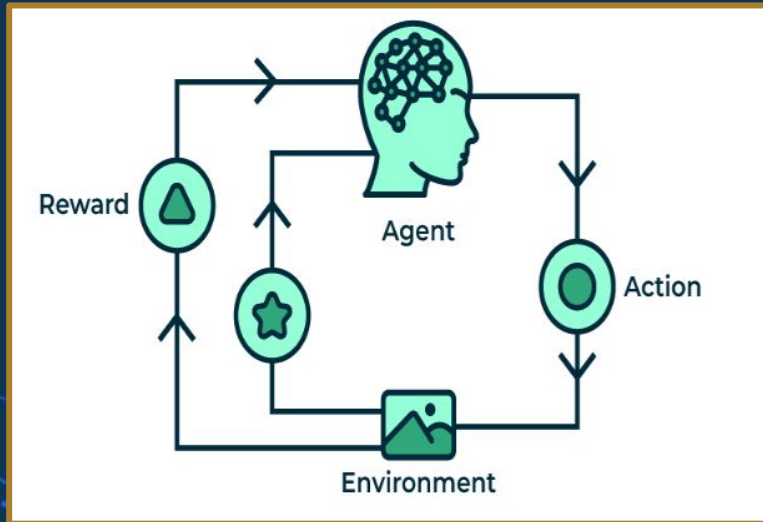
- ❖ **Unsupervised learning:** The computer learns from unlabeled data, discovering hidden patterns or structures on its own.



Source: [geeksforgeeks](https://www.geeksforgeeks.com/)

# Types of machine learning

- ❖ **Reinforcement learning:** The computer learns through interaction with an environment, receiving rewards or penalties for its actions.



Source: [geeksforgeeks](https://www.geeksforgeeks.org/)



# Some Applications of machine learning

- ❖ **Spam email filtering:** Identifying and separating spam emails from regular emails.
- ❖ **Image recognition:** Recognizing objects, faces, or scenes in images.
- ❖ **Recommender systems:** Suggesting products, movies, or songs based on user preferences.





# Supervised Learning

- ❖ In supervised learning, the algorithm learns from labelled data, which consists of input-output pairs.
- ❖ The goal is to learn a function that maps input data to the correct output labels.





# Types of Supervised Learning


- ❖ **Regression:** Predicting continuous numerical values, such as house prices or stock prices.
- ❖ **Classification:** Predicting discrete categories or classes, such as whether an email is spam or not.





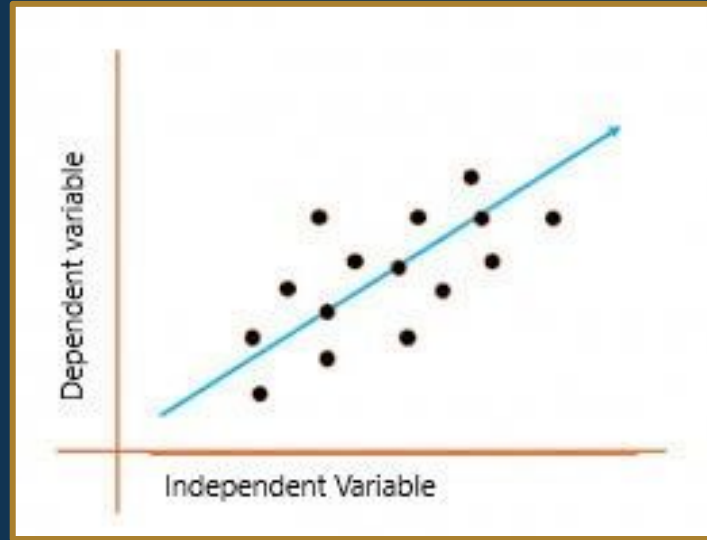


# Supervised Learning Algorithms

- ❖ **Linear regression:** Fitting a straight line to data points to make predictions.
  - ❖ **Logistic regression:** Predicting binary outcomes, such as yes/no or true/false.
  - ❖ **Decision trees:** Making decisions based on a series of questions or conditions.
  - ❖ **Support vector machines (SVM):** Finding the best boundary to separate different classes.
  - ❖ **Neural networks:** Mimicking the structure and function of the human brain to learn complex patterns.
- 

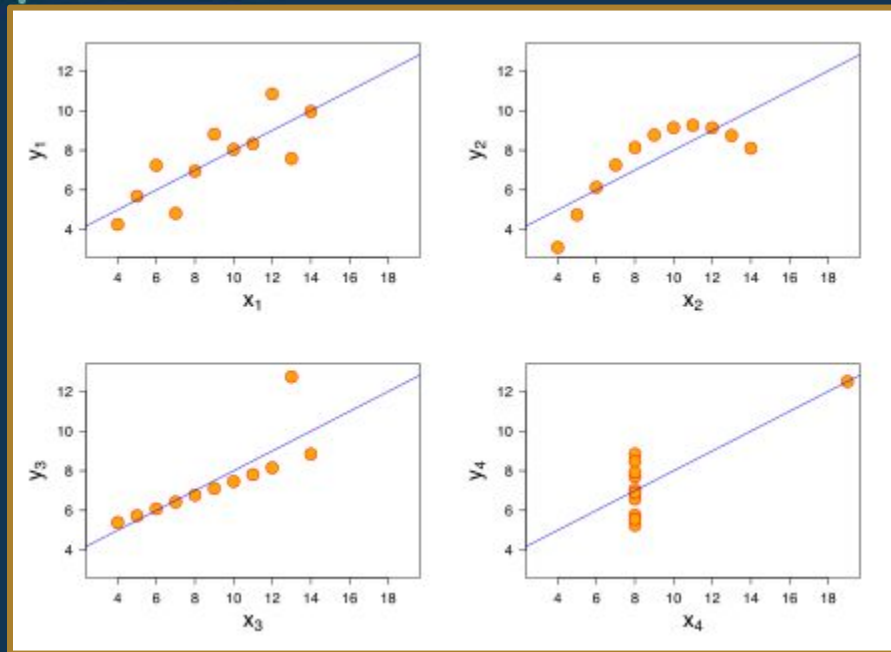
# Simple Linear Regression

- ❖ Simple linear regression is a method to study the relationship between two variables: an independent variable ( $x$ ) and a dependent variable ( $y$ ).
- ❖ It helps us understand how changes in the independent variable affect the dependent variable.



Source: [Analytics Vidhya](#)





Source: [Wikipedia](https://en.wikipedia.org/wiki/Linear_regression)



# Purpose of Simple Linear Regression

- ❖ To find the best-fitting straight line that describes the relationship between  $x$  and  $y$ .
- ❖ This line can be used to make predictions about the dependent variable based on new values of the independent variable.



# Applications of Simple Linear Regression

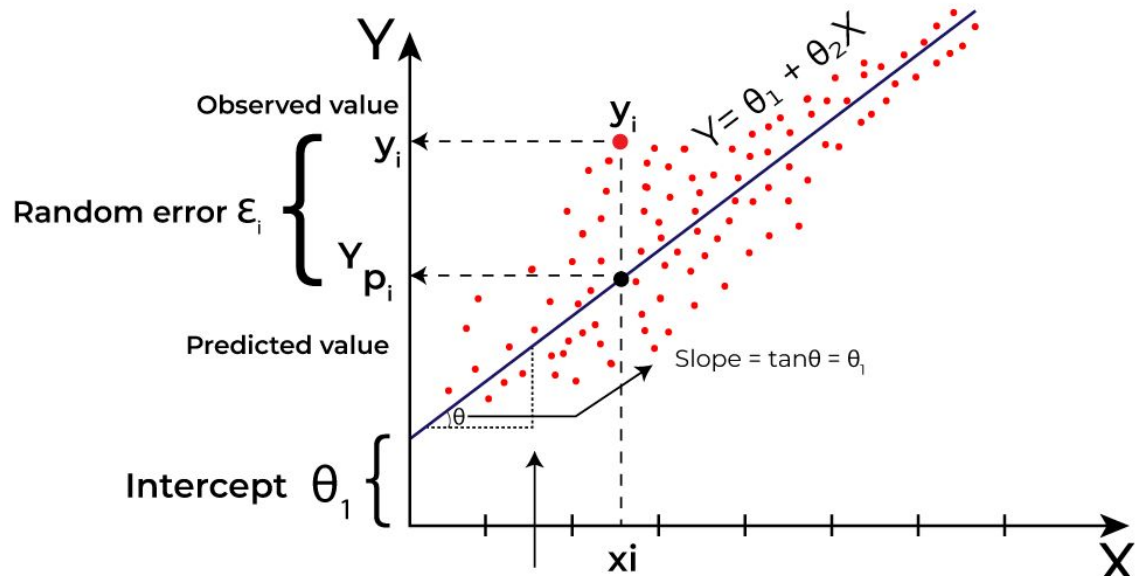
- ❖ **Sales forecasting:** Predicting future sales based on historical data.
- ❖ **Price prediction:** Estimating the price of a product based on its features.
- ❖ **Trend analysis:** Identifying trends or patterns in data over time.





# Math behind Simple Linear Regression

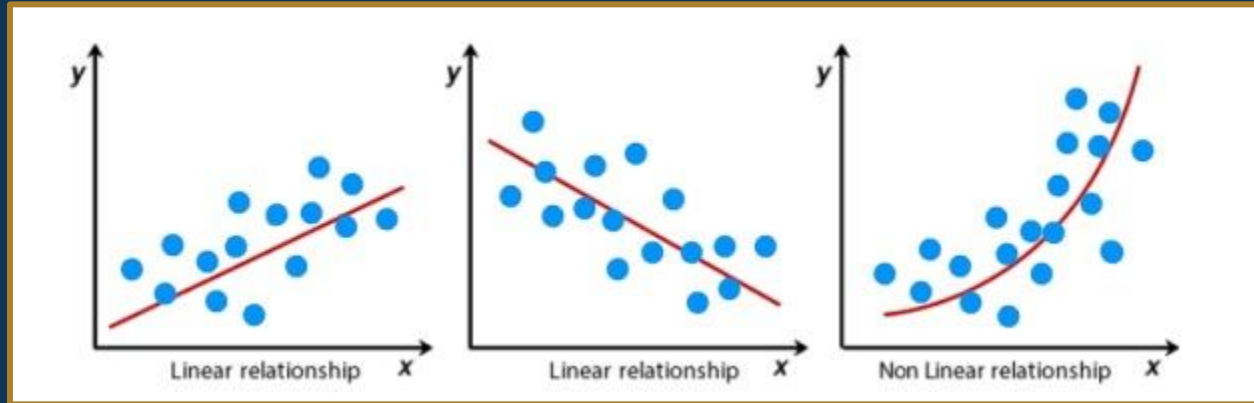
- ❖ The equation is written as:  $y = \beta_0 + \beta_1 x + \varepsilon$ 
  - $\beta_0$  is the intercept, representing the value of  $y$  when  $x$  is zero.
  - $\beta_1$  is the slope, indicating how much  $y$  changes for a one-unit increase in  $x$ .
  - $\varepsilon$  is the error term, accounting for the variability in  $y$  that cannot be explained by  $x$ .



Source: [geeksforgeeks](https://www.geeksforgeeks.org/)

# Assumptions and Limitations of Simple Linear Regression

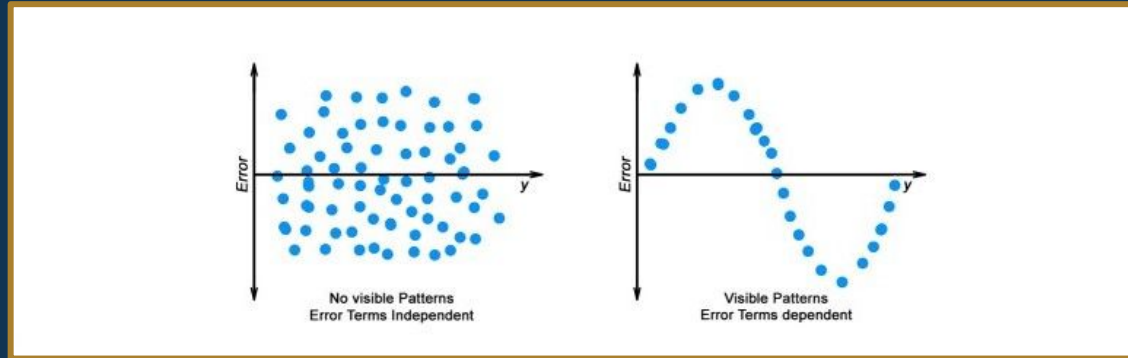
- ❖ **Linearity:** The relationship between  $x$  and  $y$  should be linear.



Source: [Analytics Vidhya](#)

# Assumptions and Limitations of Simple Linear Regression

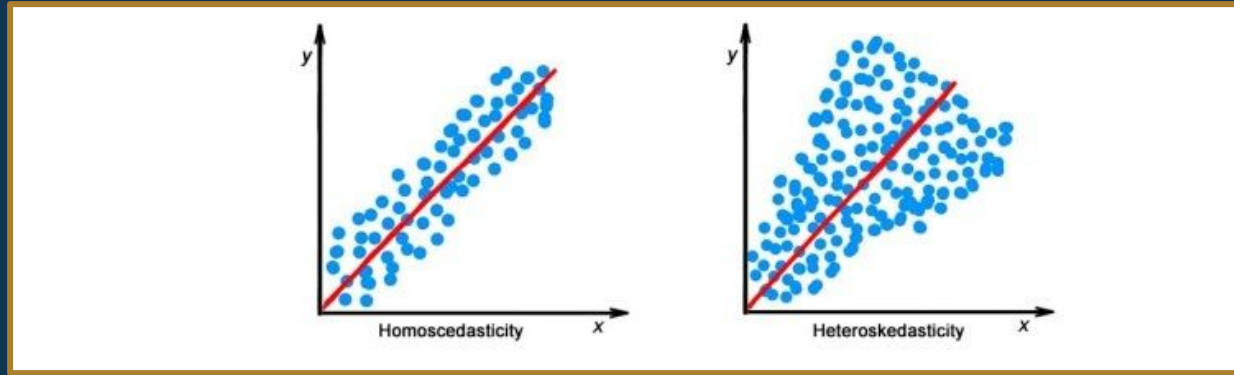
- ❖ **Independence:** The observations should be independent of each other.



Source: [Analytics Vidhya](#)

# Assumptions and Limitations of Simple Linear Regression

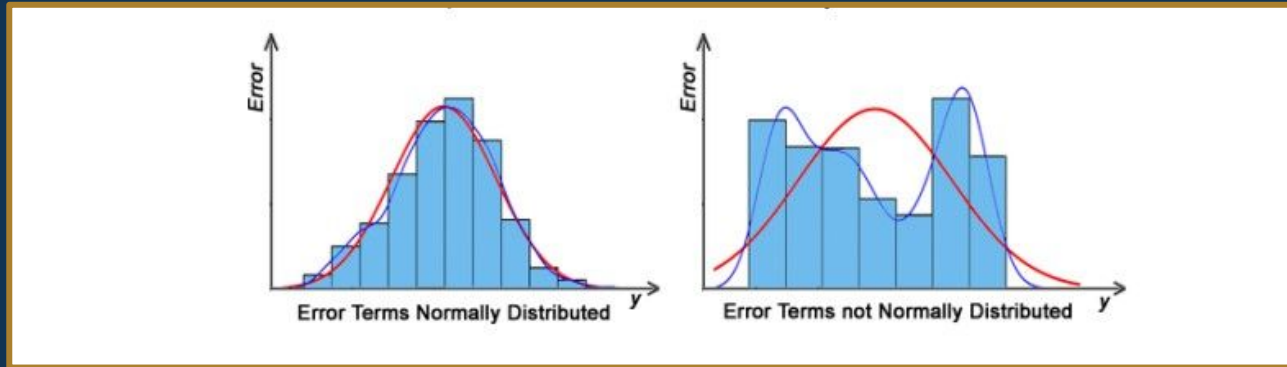
- ❖ **Homoscedasticity:** The variability of  $y$  should be constant across all values of  $x$ .



Source: [Analytics Vidhya](#)

# Assumptions and Limitations of Simple Linear Regression

- ❖ **Normality:** The errors should be normally distributed.



Source: [Analytics Vidhya](#)



# Implementing Simple Linear Regression

# Scikit-learn

- ❖ Scikit-learn is a popular Python library for machine learning.
- ❖ It provides simple and efficient tools for data analysis and modelling.

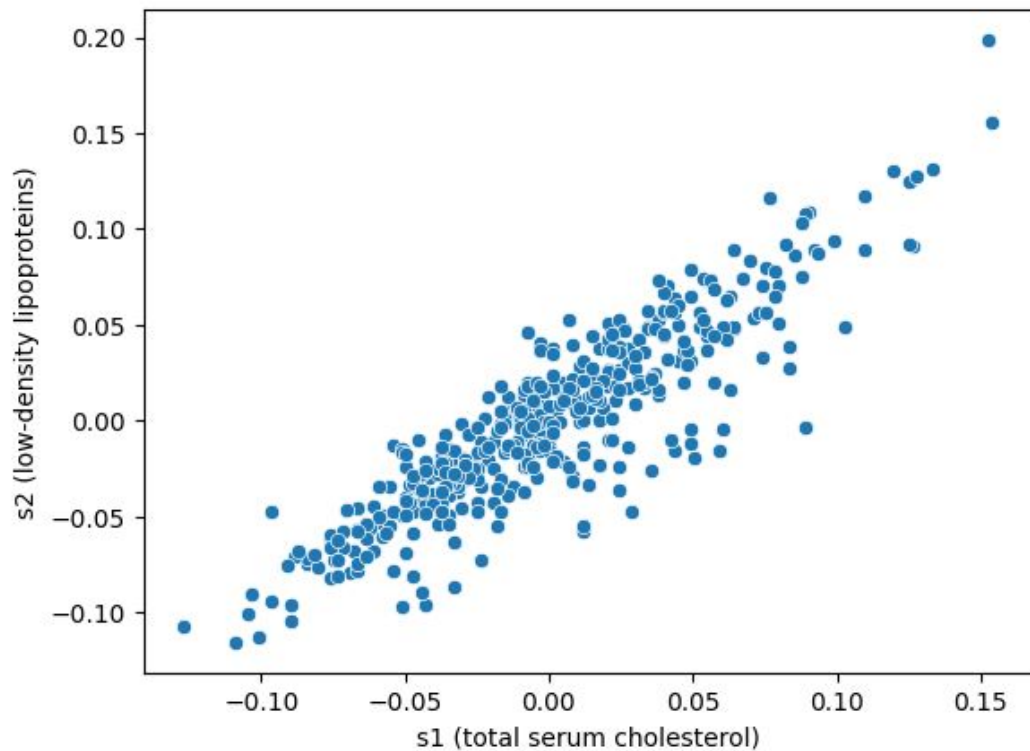
```
from sklearn.datasets import load_diabetes
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

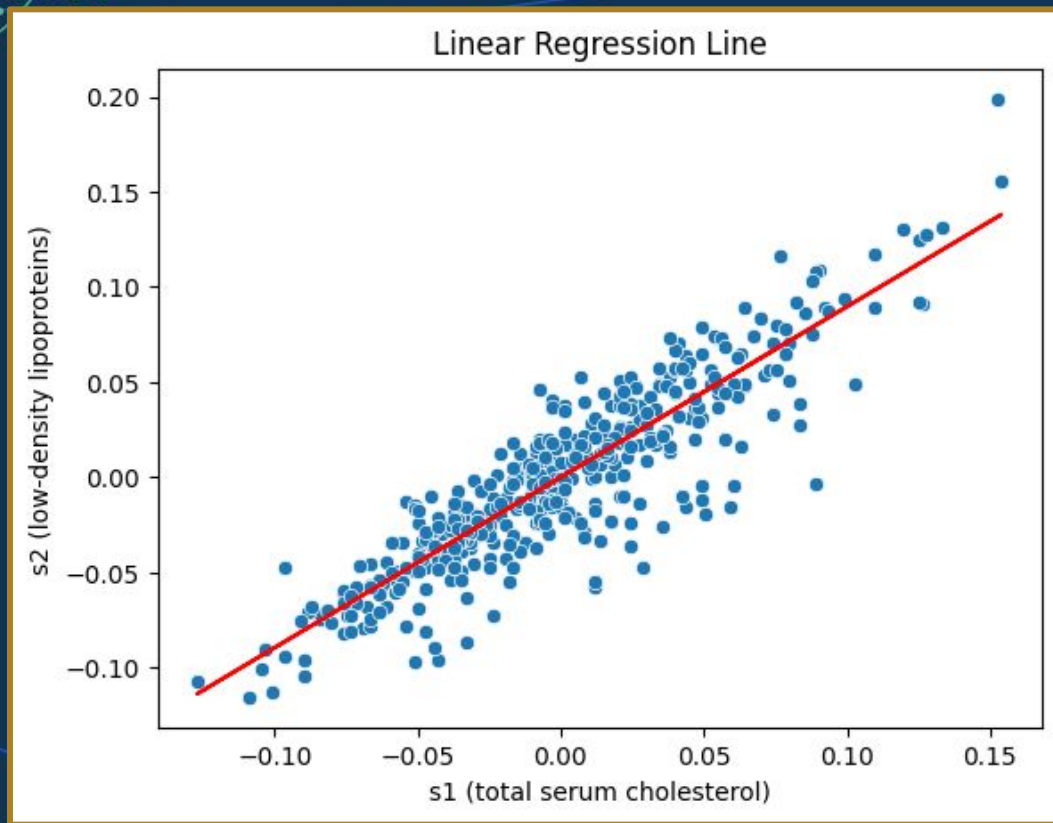
# Loading the Diabetes Dataset

- ❖ We'll use the built-in diabetes dataset from scikit-learn for our example.
- ❖ The dataset contains information about various medical predictors and a quantitative measure of disease progression.

```
df = load_diabetes(as_frame=True).data
```

Scatter Plot of s1 vs. s2







### Evaluation Metrics:

Mean Squared Error (MSE): 0.00044342882373426217

R-squared (R<sup>2</sup>) Score: 0.8040044599094562



# Interpretation of Results

- ❖ Scatter plot:
  - The scatter plot visualises the relationship between s1 (total serum cholesterol) and s2 (low-density lipoproteins).
  - It helps assess the linearity and spread of the data points.
- ❖ Linear regression line:
  - The red line represents the best-fit line obtained from the linear regression model.
  - It shows the predicted relationship between s1 and s2 based on the trained model.

# Evaluation Metrics

- ❖ Mean Squared Error (MSE):
  - MSE measures the average squared difference between the predicted and actual values.
  - A lower MSE indicates better model performance.
- ❖ R-squared ( $R^2$ ) score:
  - $R^2$  represents the proportion of variance in the target variable that can be explained by the model.
  - An  $R^2$  value closer to 1 indicates a better fit of the model to the data.

# Evaluation Metrics

❖  $R^2 = 1 - SS_{\text{error}} / SS_{\text{total}}$

❖ Where:

➤  $SS_{\text{error}} = \sum (y_i - \hat{y}_i)^2$

➤  $SS_{\text{total}} = \sum (y_i - E[y_i])^2$



# Which of the following best describes the purpose of simple linear regression?

1. To classify data points into distinct categories
2. To estimate the relationship between two continuous variables
3. To cluster data points based on their similarities
4. To predict the probability of an event occurring



# Which of the following best describes the purpose of simple linear regression?

1. To classify data points into distinct categories
- 2. To estimate the relationship between two continuous variables**
3. To cluster data points based on their similarities
4. To predict the probability of an event occurring

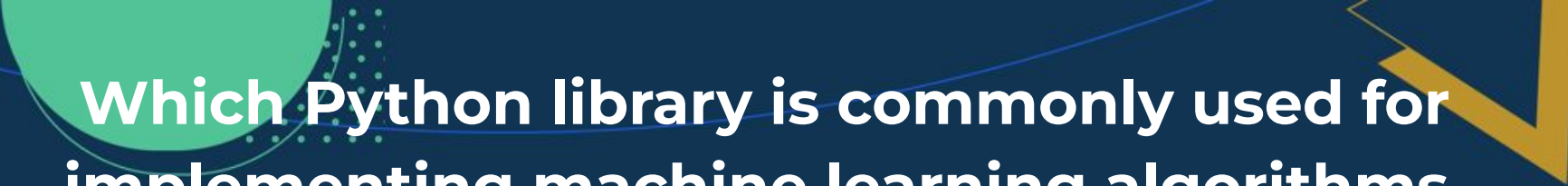
# What does the coefficient of determination (R-squared) represent in a linear regression model?

1. The average squared difference between predicted and actual values
2. The proportion of variance in the target variable explained by the model
3. The slope of the regression line
4. The intercept of the regression line

# What does the coefficient of determination (R-squared) represent in a linear regression model?

1. The average squared difference between predicted and actual values
- 2. The proportion of variance in the target variable explained by the model**
3. The slope of the regression line
4. The intercept of the regression line

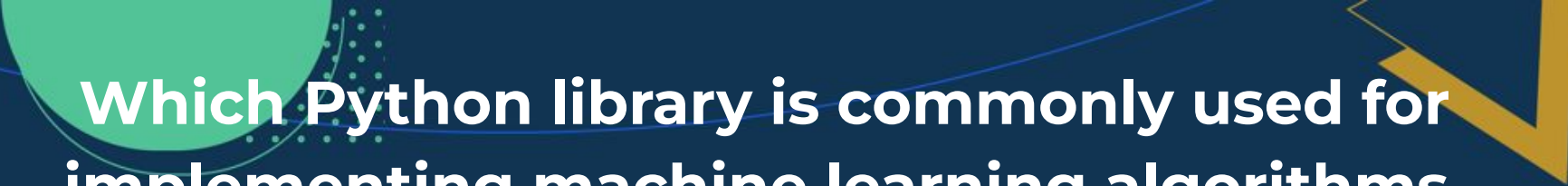




**Which Python library is commonly used for implementing machine learning algorithms, including linear regression?**

1. NumPy
2. Pandas
3. Matplotlib
4. Scikit-learn





Which Python library is commonly used for implementing machine learning algorithms, including linear regression?

1. NumPy
2. Pandas
3. Matplotlib
4. **Scikit-learn**



# What is the primary difference between supervised and unsupervised learning?

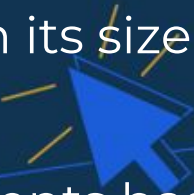
1. Supervised learning uses labelled data, while unsupervised learning uses unlabelled data
2. Supervised learning is used for classification, while unsupervised learning is used for regression
3. Supervised learning requires more computational resources than unsupervised learning
4. Supervised learning is more accurate than unsupervised learning

# What is the primary difference between supervised and unsupervised learning?

1. **Supervised learning uses labelled data, while unsupervised learning uses unlabelled data**
2. Supervised learning is used for classification, while unsupervised learning is used for regression
3. Supervised learning requires more computational resources than unsupervised learning
4. Supervised learning is more accurate than unsupervised learning

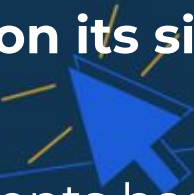


# Which of the following is an example of a regression problem?

1. Predicting the sentiment of a movie review (positive or negative)
  2. Identifying the species of a flower based on its features
  3. Estimating the price of a house based on its size and location
  4. Grouping customers into different segments based on their purchasing behaviour
- 



# Which of the following is an example of a regression problem?

1. Predicting the sentiment of a movie review (positive or negative)
  2. Identifying the species of a flower based on its features
  - 3. Estimating the price of a house based on its size and location**
  4. Grouping customers into different segments based on their purchasing behaviour
- 



# Questions and Answers





# Thank you for attending



Department  
for Education

CoGrammar

