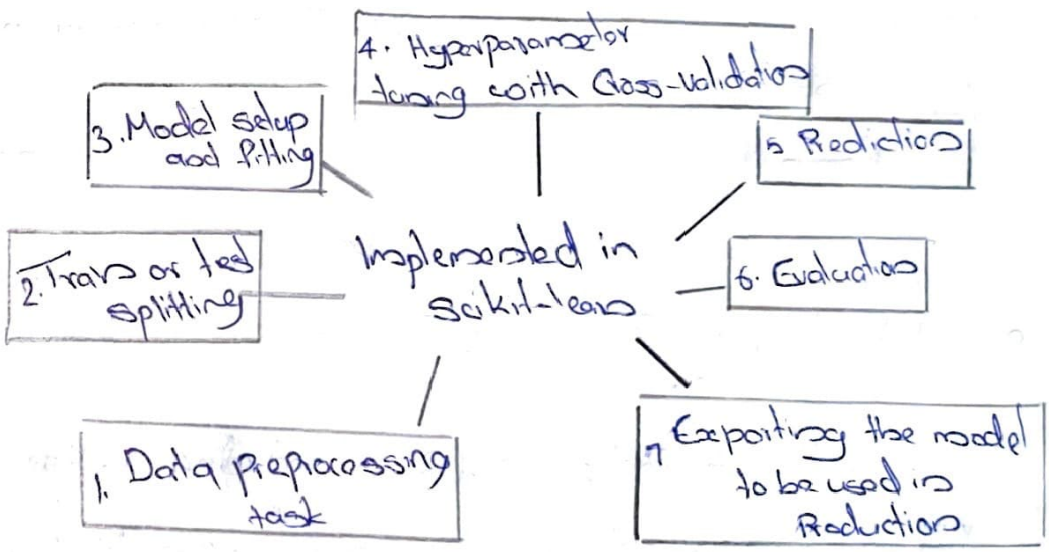
- Scikit-learn
  - Free ML Library for Python
  - Classifications, regression, clustering and dimensionality reduction algorithm
  - Designed to work with NumPy and Scipy
  - Excellent documentation and community support
  - Constantly evolving
  - Enables easy implementation of ML models

- Machine learning Pipeline tasks



Module 2

- Linear Regression
  - Supervised Learning model
  - Models a relationship between a continuous target variable and explanatory features

- Types of regression
  - Simple regression (Simple linear, Simple Nonlinear)
  - Multiple regression (Multiple linear, Multiple Nonlinear)

- Regression algorithms
  - Linear and polynomial
  - Random forest
  - Extreme Gradient Boosting (XGBoost)
  - K-nearest neighbors (KNN)
  - Support vector Machines (SVM)
  - Neural Networks

Simple Linear Regression (SLR)

The SLR is a statistical technique in machine learning that fits a linear equation to Predict the output (dependent variable) from a single input (independent variable). It assumes a straight-line relationship between input and Output.

Example:

$$y = \text{intercept} + \text{slope} * x$$

- Multiple Linear Regression (MLR)

MLR generalizes SLR by including multiple input features. It used when the outcome depends on several factors. The model fits a hyperplane in n-dimensional space to minimize the error between predicted and actual outcomes.

$$y = \text{intercept} + (\text{coef}_1 * x_1) + (\text{coef}_2 * x_2) + \ldots + (\text{coef}_n * x_n)$$

- Polynomial regression

  - Relationship between independent variable X and the dependent variable y is modelled as an nth degree polynomial in X

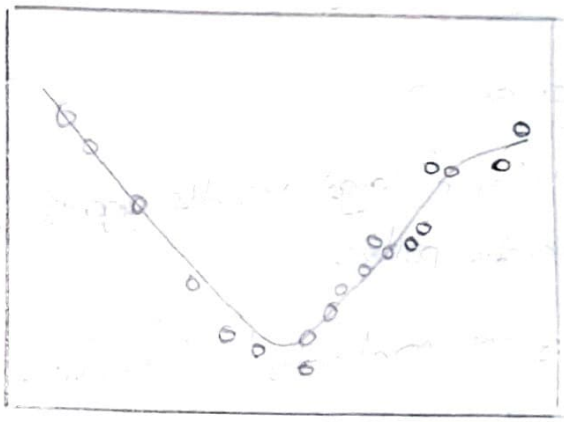
  - Equation (degree 2)

$$y = \theta_0 + \theta_1 x + \theta_2 x^2$$

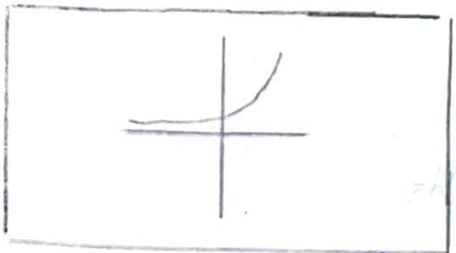
  - Overfitting:

- Polynomial regression model 'memorizes' everything including noise or variation
- Pick a regression that fits data without overfitting

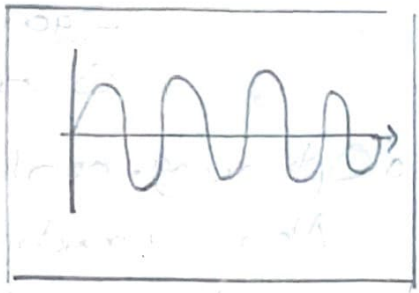
- Application of non linear regression.



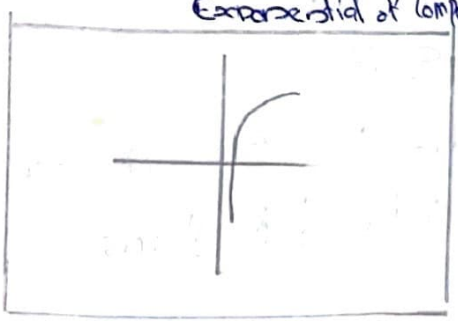
- Polynomial regression
  - Non linear dependence on input features but linear dependence on regression coefficients
  - Can be transformed into a linear regression problem
  - Real-world, complex, non linear relationship can't be modeled as polynomial



Exponential of Compound growth



Periodicity



Logarithmic

- Compound growth example (Exponential)
  - Scatterplot displays strong dependence of GDP on time, but relationship is nonlinear
  - GDP increasing overtime, but growth rate is also increasing

$$\hat{y} = \theta_0 + \theta_1 e^x$$

- Productivity by work hours example (Logarithmic)
  - Work more hour per day on average increases Productivity
  - After a reasonable limit, each additional hour generates less Productivity

---

① Linear or nonlinear regression

- Analyzing scatterplots of target variable against input variable reveals patterns
- Express patterns as mathematical functions and determine if:
  - linear
  - Exponential
  - Logarithmic
  - Sinusoidal

- Optimizing nonlinear models.

Nonlinear machine learning models.

- Regression trees
- Random forest
- Neural Networks
- SVM
- Gradient Boosting Machine
- K-Nearest Neighbors