### **Week 8: Introduction to Machine Learning**

#### **Overview**

This week, we will introduce you to the basics of machine learning, a powerful tool for analyzing data and making predictions. We will cover the fundamental concepts of machine learning, including supervised and unsupervised learning, and explore simple linear regression and classification using Scikit-Learn, a popular machine learning library in Python. By the end of this week, you will have a foundational understanding of machine learning and be able to implement basic models using Python.

#### **Learning Objectives**

By the end of this module, students will be able to:

1. Understand the basics of machine learning and its applications.
2. Differentiate between supervised and unsupervised learning.
3. Implement simple linear regression using Scikit-Learn.
4. Perform basic classification tasks using Scikit-Learn.

#### **Basics of Machine Learning**

**Machine Learning**: A subset of artificial intelligence that involves the use of algorithms and statistical models to enable computers to improve their performance on a task through experience (data).

**Applications**: Machine learning is used in various fields, including:

* Predictive analytics
* Natural language processing
* Image and speech recognition
* Recommender systems
* Autonomous vehicles

#### **Types of Machine Learning**

1. **Supervised Learning**: The algorithm is trained on labeled data, meaning that each training example is paired with an output label. The goal is to learn a mapping from inputs to outputs.
   * **Regression**: Predicts continuous values (e.g., predicting house prices).
   * **Classification**: Predicts discrete values (e.g., identifying spam emails).
2. **Unsupervised Learning**: The algorithm is used on data without labels, and the goal is to find patterns or structures within the data.
   * **Clustering**: Grouping similar data points together (e.g., customer segmentation).
   * **Dimensionality Reduction**: Reducing the number of random variables under consideration (e.g., PCA).

#### **Introduction to Scikit-Learn**

Scikit-Learn is a robust and versatile machine learning library in Python. It provides simple and efficient tools for data mining and data analysis, and is built on NumPy, SciPy, and Matplotlib.

**Installing Scikit-Learn**:

pip install scikit-learn

#### **Implementing Simple Linear Regression**

Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things:

1. Does a set of predictor variables do a good job in predicting an outcome (dependent) variable?
2. Which variables in particular are significant predictors of the outcome variable, and in what way do they impact the outcome variable?

**Example**:

import numpy as np

import matplotlib.pyplot as plt

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

# Sample data

X = np.array([[1], [2], [3], [4], [5]])

y = np.array([1, 4, 9, 16, 25])

# Split data into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=0)

# Create and train the model

model = LinearRegression()

model.fit(X\_train, y\_train)

# Predict using the model

y\_pred = model.predict(X\_test)

# Plot results

plt.scatter(X, y, color='blue')

plt.plot(X\_test, y\_pred, color='red')

plt.title('Linear Regression Example')

plt.xlabel('X')

plt.ylabel('y')

plt.show()

#### **Performing Basic Classification**

Classification is a supervised learning technique used to predict the category of a data point. Examples include email spam detection, sentiment analysis, and object recognition.

**Example**:

from sklearn.datasets import load\_iris

from sklearn.model\_selection import train\_test\_split

from sklearn.neighbors import KNeighborsClassifier

from sklearn.metrics import accuracy\_score

# Load dataset

iris = load\_iris()

X = iris.data

y = iris.target

# Split data into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=0)

# Create and train the model

model = KNeighborsClassifier(n\_neighbors=3)

model.fit(X\_train, y\_train)

# Predict using the model

y\_pred = model.predict(X\_test)

# Evaluate the model

accuracy = accuracy\_score(y\_test, y\_pred)

print(f'Accuracy: {accuracy:.2f}')

#### **Learning Activities**

To reinforce your understanding of this week's content, complete the following activities:

1. **Reading Assignment**: Read Chapter 8 of "Data Analytics Made Accessible" by Anil Maheshwari, focusing on machine learning basics.
2. **Video Lecture**: Watch the video "Introduction to Machine Learning with Python" on YouTube to see practical examples of machine learning.
3. **Hands-On Exercise**: Implement a simple linear regression and a classification model using Scikit-Learn with a sample dataset in Jupyter Notebook.

#### **Discussion Questions**

Participate in the class discussion by answering the following questions:

1. What are the differences between supervised and unsupervised learning?
2. How can linear regression be used in real-world scenarios?
3. What challenges might you face when implementing machine learning models?

#### **Summary**

This week, we explored the basics of machine learning, including the differences between supervised and unsupervised learning. We implemented simple linear regression and classification models using Scikit-Learn, a powerful machine learning library in Python. These foundational skills are essential for any data analyst to analyze data and make predictions effectively.

#### **Additional Resources**

* **Book**: "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron.
* **Website**: Visit Scikit-Learn Documentation for comprehensive guides on using Scikit-Learn for machine learning.
* **Tutorial**: Follow the tutorial "Introduction to Machine Learning with Python" on [Kaggle](https://www.kaggle.com/) to practice with real datasets.

#### **Homework**

1. Write a short essay (300-500 words) on the potential applications of machine learning in a field of your choice.
2. Complete the hands-on exercise and submit your Jupyter Notebook with the machine learning models through the course portal.