Day37_Introduction_to_Machine_Learning

July 5, 2025

Day 34 – Introduction to Machine Learning (ML)

In this notebook, we explore the basics of ML through a simple example of Supervised Learning using Linear Regression.

1 What is Machine Learning?

Machine Learning (ML) is a subset of Artificial Intelligence (AI) that enables systems to learn patterns from data and make decisions without being explicitly programmed. Instead of writing rules manually, ML algorithms find patterns in data and use those patterns to make decisions or predictions.

Machine Learning is used in many everyday applications like:

- Email spam detection
- Product recommendations
- Voice assistants
- Stock price prediction

1.1 Traditional Programming vs Machine Learning

Traditional Programming	Machine Learning
Input + Logic = Output	Input + Output = Learn Logic (Model)
Human-written rules	Machine learns rules from data

1.2 Structured vs Unstructured Data

Structured Data	Unstructured Data
Stored in Excel, CSV, Databases	Emails, Images, Videos, Audio
Applied in ML	Used in Deep Learning or NLP

1.3 Three Phases in ML Workflow

1. **Training Phase** – Learn from data
2. **Validation Phase** – Tune model parameters
3. **Testing Phase** – Evaluate model accuracy

2 Types of Machine Learning

2.1 Supervised Learning

- **Regression** (predict numbers)
- **Classification** (predict categories)

```python
[4]: # Example: Supervised Learning (Regression)

from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
import numpy as np

# Larger dataset
X = np.array([[1], [2], [3], [4], [5], [6], [7], [8]])
y = np.array([2, 4, 6, 8, 10, 12, 14, 16])  # y = 2 * x

# Split into train and test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4)

# Train model
model = LinearRegression()
model.fit(X_train, y_train)

# Predict
prediction = model.predict(X_test)

# Show results
print("Test Inputs (X_test):\n", X_test)
print("Predicted Outputs:\n", prediction)
print("Actual Outputs:\n", y_test)
```

```
Test Inputs (X_test):
 [[2]
 [5]
 [6]
 [8]]
Predicted Outputs:
 [ 4. 10. 12. 16.]
Actual Outputs:
 [ 4 10 12 16]
```

2.2 Unsupervised Learning

- **Clustering** (e.g., K-Means)

```python
[2]: # Example: Unsupervised Learning (Clustering)
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
```

```
X = np.array([[1, 2], [2, 3], [10, 12], [11, 14], [5, 5]])

kmeans = KMeans(n_clusters=2)
kmeans.fit(X)
labels = kmeans.labels_

plt.scatter(X[:, 0], X[:, 1], c=labels)
plt.title("K-Means Clustering")
plt.show()
```

C:\Users\Lenovo\anaconda3\Lib\site-packages\sklearn\cluster_kmeans.py:1419:
UserWarning: KMeans is known to have a memory leak on Windows with MKL, when
there are less chunks than available threads. You can avoid it by setting the
environment variable OMP_NUM_THREADS=1.
  warnings.warn(
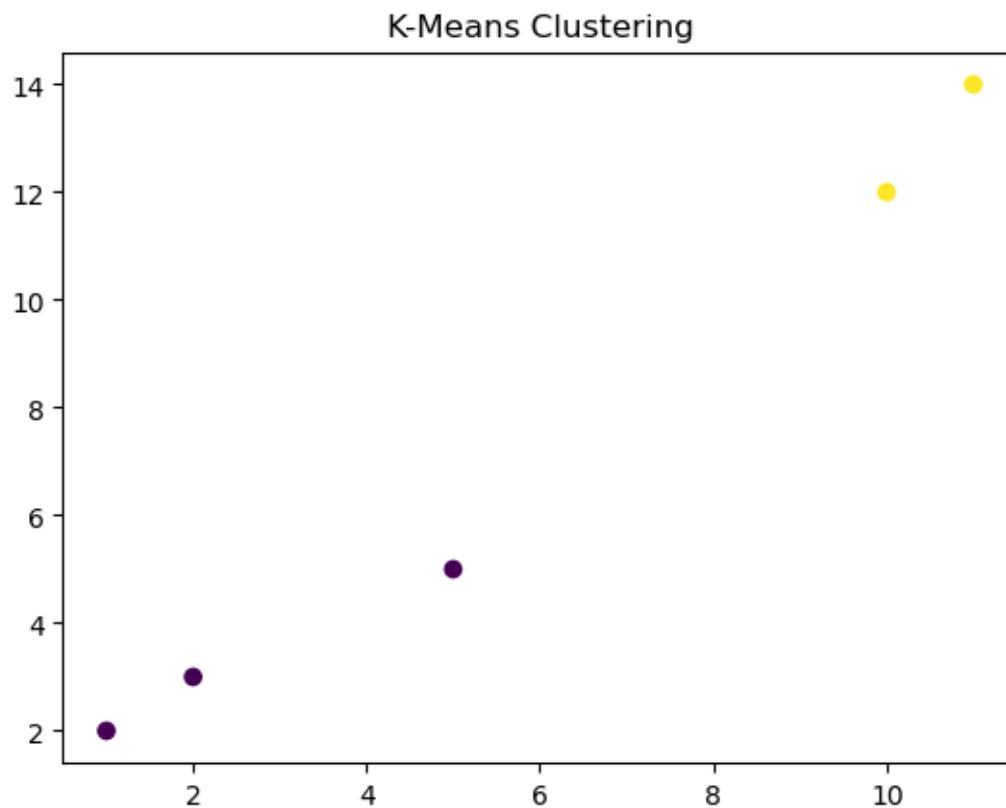
3   7 Steps of a Machine Learning Project

1. **Gather Data**
2. **Prepare Data**

3. **Choose Model**
4. **Train Model**
5. **Evaluate Model**
6. **Tune Parameters (Hyperparameter Tuning)**
7. **Make Predictions**

4 Train-Test Split Example

```python
from sklearn.model_selection import train_test_split
import pandas as pd

# Example Dataset
data = {'Hours': [1, 2, 3, 4, 5], 'Scores': [20, 40, 60, 80, 100]}
df = pd.DataFrame(data)

X = df[['Hours']]  # independent
y = df['Scores']   # dependent

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

print("Training X:\n", X_train)
print("Testing X:\n", X_test)
```

```
Training X:
    Hours
4      5
2      3
3      4
1      2
Testing X:
    Hours
0      1
```

5 Tools & IDEs for ML

Tool	Use
Jupyter Notebook	Data Analysis and Documentation
VS Code	Powerful IDE for ML Model Building
Google Colab	Free Cloud Notebook

6 Key Takeaways

- Machine Learning uses **data to learn logic** (rules)
- Two main types: **Supervised** and **Unsupervised**
- ML workflow involves **data prep, model training, and evaluation**
- Tools like **scikit-learn**, **Pandas**, and **Matplotlib** help implement ML in Python