

General Python Coding Concepts

Control Structures

- if-else statements.
- elif for multiple conditions.

Examples:

if condition:

 # Do something

else:

 # Do something else

if condition1:

 # Do something

elif condition2:

 # Do something else

Loops and Iteration

- Using for loops with lists and enumerate().

for item in iterable:

 # Do something with each item

for index, value in enumerate(my_list):

 # Do something with the index and value

Common Python Functions and Libraries

- Basic list operations.
- Importing libraries like numpy, pandas, and scikit-learn.

Python and pandas

pandas Basics

- Data Structures:
 - Series: A one-dimensional array-like object.
 - DataFrame: A two-dimensional, tabular data structure.
- Accessing Data:
 - `.loc[]`: Used for label-based indexing (access by row labels and column names).
 - `.iloc[]`: Used for positional indexing (access by row and column indices).
- Applying Functions:
 - `.apply()`: Apply a function along an axis of a DataFrame (row-wise or column-wise).
- Viewing Data:
 - `.head()`: View the first 5 rows of a DataFrame (default).
 - `.tail()`: View the last 5 rows of a DataFrame.

File Handling in pandas

- Reading CSV Files:
 - `pd.read_csv()`: Function to read CSV files and convert them into DataFrame objects.

Data Cleaning in pandas

- Handling Messy Data:
 - Renaming columns, dropping missing values, handling duplicates, etc.

Data Visualization with Seaborn and Matplotlib

Seaborn Basics

- What is Seaborn?
 - Seaborn is built on top of Matplotlib and provides a high-level interface for creating attractive and informative statistical graphics.
- Key Plot Types in Seaborn:
 - `scatterplot()`: To visualize relationships between two variables.
 - `lineplot()`: To visualize trends over time.
 - `barplot()`: To compare categorical data.
 - `histplot()`: To plot histograms.

Matplotlib Basics

- What is Matplotlib?
 - Matplotlib is a lower-level library used to create static, animated, and interactive plots in Python.
- Basic Plotting:

- `plot()`: Create line plots.
- `scatter()`: Create scatter plots.
- `hist()`: Create histograms.

Combining Seaborn and Matplotlib

- Using Seaborn with Matplotlib:

Seaborn can be used to create plots that are customized with Matplotlib. You can adjust the layout, labels, and other elements using Matplotlib functions.

```
import seaborn as sns
import matplotlib.pyplot as plt
sns.set(style='whitegrid') # Customizing Seaborn style
sns.barplot(x='feature', y='target', data=my_data)
plt.title('My Seaborn Plot')
plt.show()
```

Object-Oriented Programming in Python (Classes)

Classes in Python

- Definition: A class is a blueprint for creating objects that share common properties and methods.
- Basic Structure:
 - Define a class using the `class` keyword.
 - Use `__init__()` to initialize instance attributes.
 - Define methods that operate on the attributes.

```
class Person:
    def __init__(self, name, age):
        self.name = name
        self.age = age

    def greet(self):
        return f"Hello, my name is {self.name} and I am {self.age} years old."

# Create an object
person1 = Person("Alice", 30)
print(person1.greet())
```

Inheritance:

- Allows a new class to inherit properties and methods from an existing class

Linear Regression

Basics of Linear Regression

- Purpose: Linear regression predicts a continuous target variable by modeling the relationship between the dependent variable (Y) and independent variables (X).

Key Metrics for Evaluation

- Mean Squared Error (MSE): Measures the average of the squared differences between predicted and actual values.
- R-squared: Represents how well the model fits the data (a value between 0 and 1, where 1 means perfect fit).

Methods to Know:

- `fit()`: Trains the model on input data.
- `predict()`: Makes predictions based on input features.
- `score()`: Returns the R-squared value of the model.

K-Nearest Neighbors (KNN)

KNN Basics

- **Definition and Purpose of KNN**
 - KNN is a supervised learning algorithm used for classification and regression. It makes predictions based on the majority class or average value of the k-nearest neighbors in the feature space.
- **Difference Between Supervised and Unsupervised Learning**
 - Supervised learning involves learning from labeled data (e.g., KNN), where the model is trained on input-output pairs.
 - Unsupervised learning involves data without labels, where the model tries to find hidden patterns (e.g., clustering).

Key Parameters in KNN

- `n_neighbors`: Selecting the number of neighbors. Commonly, odd numbers are chosen to avoid ties.
- `metric`: Common distance metrics include Euclidean (straight-line distance) and Manhattan (sum of absolute differences).
- `weights`:
 - Uniform: All neighbors are given equal weight.
 - Distance: Neighbors closer to the query point are given more weight.

Advantages and Disadvantages of KNN

- Advantages:
 - Simple and intuitive.
 - No training phase (instance-based learning).
 - Works well with small datasets.
- Disadvantages:
 - Sensitive to noisy data.
 - Computationally expensive with large datasets.
 - High memory usage since it stores the entire dataset.