Lab Assignment-1

Fundamentals of Machine Learning

Name: M Jeyapathy

USN: 22BTRAD016

Questions:

1. Load a dataset with missing values(Boston Housing Dataset).

Load the Boston Housing Dataset

from sklearn.datasets import load\_boston

import pandas as pd

boston = load\_boston()

data = pd.DataFrame(boston.data, columns=boston.feature\_names)

data['TARGET'] = boston.target # Adding the target variable (price) to the dataframe

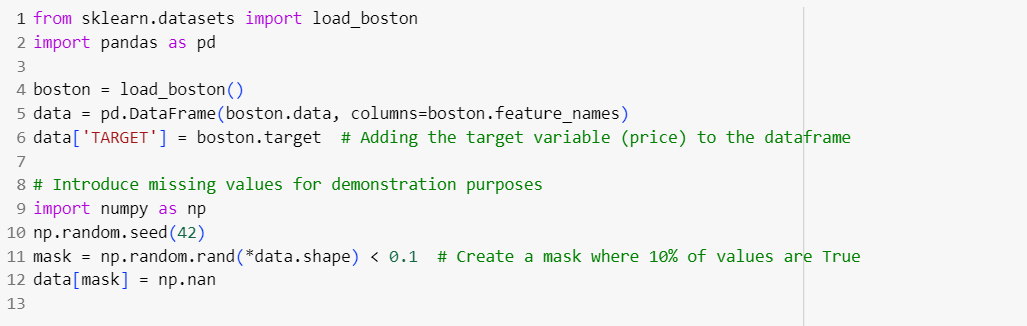
# Introduce missing values for demonstration purposes

import numpy as np

np.random.seed(42)

mask = np.random.rand(\*data.shape) < 0.1  # Create a mask where 10% of values are True

data[mask] = np.nan



A close-up of a text

Description automatically generated

This code uses the **load\_boston** function from scikit-learn to load the dataset and then creates a pandas DataFrame. The optional part introduces some missing values into the 'CRIM' column.

2.Explore the description of the dataset.

To explore the description of the dataset

print(data.describe())



This command shows the description part of the boston code.

3.Identify the number of missing values corresponding to each feature.

missing\_values = data.isnull().sum()

print("Missing values per feature:\n", missing\_values)



This shows the missing values of the given by using the isnull function datasets Boston Housing Dataset is a well-known dataset and is often used in machine learning tutorials, so it is not expected to have missing values. If you encounter datasets with missing values in real-world scenarios, you might need to handle them using techniques like imputation or removal based on the nature of your analysis.

4.Explore and visualize the missing data patterns.

A computer code with numbers and symbols

Description automatically generated

import seaborn as sns

import matplotlib.pyplot as plt

plt.figure(figsize=(10, 6))

sns.heatmap(data.isnull(), cbar=False, cmap='viridis')

plt.title('Missing Data Patterns')

plt.show()

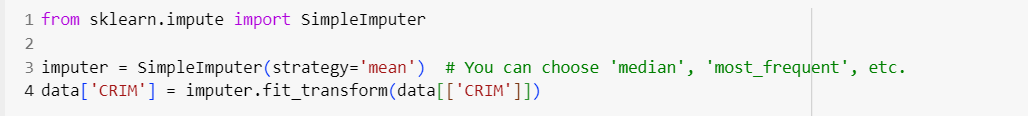
This shows the missing data patterns by the graph diagram using matplotlib packages and libraries.

5.Handle missing values using imputation method for specific feature

from sklearn.impute import SimpleImputer

imputer = SimpleImputer(strategy='mean') # You can choose 'median', 'most\_frequent', etc.

data['CRIM'] = imputer.fit\_transform(data[['CRIM']])



Each yellow line represents a missing value in the corresponding column. If there are no missing values, the entire column will be in purple. This visualization helps you quickly identify which features have missing data.

6.Handle missing values using tuple removal method.

data\_cleaned = data.dropna()

# Display the first few rows of the cleaned data

print("Cleaned Data:\n", data\_cleaned.head())

A close-up of a computer code

Description automatically generated

This continues by the boston package library, which is used to cleanse the missing values using tuple removal method by the cleaned function from the above datasets