Practical-2

Introduction to reproducible Machine Learning Operations

The aim of the practial is to get the hands-on experience of reproducing the machine learning operations at each stage. Student needs to apply the following steps in the practical.

Step 1 :

Ensure that the numpy, scikit learn, and matplotlib libraries are available in your system. Create the requirements.txt file and make a note of the versions of these libraries.

To check the Versions of Libraries :

```python
import numpy as np
print(np.__version__)

1.23.5


import sklearn
print(sklearn.__version__)

1.2.2


import matplotlib
print(matplotlib.__version__)

3.7.1


import pandas
print(pandas.__version__)

1.5.3
```

Step 2 :

    a) Import the data and scale it using StandardScaler:

```
import numpy as np
from sklearn.preprocessing import StandardScaler
import joblib

data = np.genfromtxt('/content/Sample.txt', delimiter=',')

scaler = StandardScaler()
scaled_data = scaler.fit_transform(data)

joblib.dump(scaler, 'scaler_object.joblib')
```

    b) Splitting the normalized data:

```
from sklearn.model_selection import train_test_split

# Split the data into training and testing sets
train_data, test_data = train_test_split(scaled_data, test_size=0.2, random_state=42)
```

    c) Storing the snapshot of the data as a numpy file:

```
np.save('train_data.npy', train_data)
np.save('test_data.npy', test_data)
```

Step 3 :

Apply the linear regression algorithm on the dataset and assess the prediction on the test dataset.

a) Store the trained model into the local file system to ensure the reproducibility of the prediction. Import the model and the test dataset into other python file. Check whether the same prediction is obtained in the latter case.

```
from sklearn.linear_model import LinearRegression
import joblib

X_train = train_data[:, :-1]
y_train = train_data[:, -1]

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

X_test = test_data[:, :-1]
y_test = test_data[:, -1]

predictions = model.predict(X_test)

joblib.dump(model, 'linear_regression_model.joblib')
```
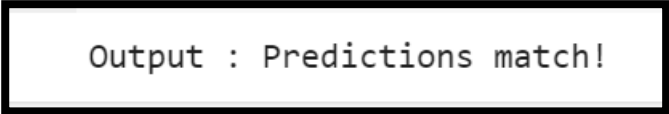
```
loaded_model = joblib.load('linear_regression_model.joblib')
loaded_predictions = loaded_model.predict(X_test)

if np.array_equal(predictions, loaded_predictions):
    print("Output : Predictions match!")
else:
    print("Output : Predictions differ!")
```

Output :

```
Output : Predictions match!
```