Machine Learning Techniques KAI-651

PRACTICAL-01

Experiment-1 [Simple Linear Regression]

Aim: Create a Linear Regression model which predicts target value based on input features after training on given dataset.

Dataset: Fuel Consumption CO_2.csv

Objective: Use Scikit Learn
Create a model, train it, test it.

Download dataset fuel Consumption CO_2.csv
Import matplotlib and pandas
→ Load the dataset file into a pandas dataframe using appropriate pandas f°(read.csv)

After loading the dataset into the dataframes (df). The head() function is used to display the first few rows of data frame.

Date Exploration :- It involves examining, visualising and summarizing the data to gain insights into its structure, pattern and relationships between variables.
In this CO_2 emissions varies almost linearly. So, we check enginize with emission.

Train and Test Split : It is a technique used to evaluate the performance of a machine learning model.
→ Here we splits the data into two subsets training set and test set.

Teacher's Signature : _____

```
import matplotlib.pyplot as plt
import pandas as pd
df = pd.read_csv(r'/content/fuel/consumptionCO2.csv')
df.head()
plt.scatter(df.FUELCONSUMPTION_COMB, df.CO2EMISSIONS, color='blue')
plt.xlabel("FUELCONSUMPTION_COMB")
plt.ylabel("EMISSION")
plt.show()
from sklearn.model_selection import train_test_split train_set, test_set=
 train_test_split(df, test_size=0.2, random_state=42)
df_train = train_set.copy()
df_train.head()
from sklearn import linear_model
regr = linear_model.LinearRegression()
x_train = df_train[['ENGINE SIZE']];
y_train = df_train[['CO2EMISSIONS']];
reg.fit(x_train, y_train)
print('coefficients:', regr_intercept)
print('Intercept:', regr_coef)
plt.scatter(df_train.ENGINESIZE, df_train.CO2EMISSIONS, color='blue')
plt.plot(x_train, regr.coef_[0][0]* x_train + regr.intercept_[0], '-r']
plt.xlabel("Engine size")
plt.ylabel("Emission")
from sklearn.metrics import r2_score
df_test = test_set.copy()
x_test = df_test[['ENGINESIZE']];
y_test = df_test[['CO2EMISSIONS']];
y_predict = regr.predict(x_test);
r_square = r2_score(y_test, y_predict)
print(r_square)
plt scatter(x_test, y_test, color='blue')
plt scatter(x_test, y_predict, color='red')
```

Date 27-02-2024

Expt. No. 1 Page No. 2

Train a model : It is a process of fitting a mathematical algorithm to a dataset in order to learn patterns and relationships.

→ Model predict CO_2 emission gives one feature 'Engine size'.

→ Through this we get two weights :- coefficients [126.28970217] and Intercept [38.99297872]

- Evaluation : It is a process of assessing the performance of a trained model on unseen data.

→ It involves comparing the model's predictions with the actual target in the test dataset.

- Visualize our predicted values

→ It is a way to understand how well your model is performing and to gain insights into its behaviour.

Conclusion : The value of r-square for this dataset is 0.7615597

Teacher's Signature :