Linear Regression - Explainability

Dr. Sabu M K
Professor
Department of Computer Applications
CUSAT

- import pandas as pd
- import numpy as np
- from sklearn import model_selection
- from sklearn.linear_model import LinearRegression
- from sklearn.model_selection import train_test_split

- # Step 2 - Reading the Data and Performing Basic Data Checks
-
- df = pd.read_csv('/home/tech/Desktop/regressionexample.csv')
- print(df.shape)
- df.describe()

- target_column = ['unemploy']
- predictors = list(set(list(df.columns))-set(target_column))
- df[predictors] = df[predictors]/df[predictors].max()
- df.describe()

- #Step 4 - Creating the Training and Test Datasets
-
- X = df[predictors]
- y = df[target_column]
-
- X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30, random_state=40)
- print(X_train.shape); print(X_test.shape)

- X_train.head()

- model = LinearRegression()
-
- model.fit(X_train, y_train)
-
- print(f"model score on training data: {model.score(X_train, y_train)}")
- print(f"model score on testing data: {model.score(X_test, y_test)}")

```
import matplotlib.pyplot as plt

coefs = pd.DataFrame(
    model.coef_.reshape(4,1), columns=["Coefficients"], index = X_train.columns
)

coefs.plot(kind="barh", figsize=(9, 7))
plt.title("Linear Regression model")
plt.axvline(x=0, color=".5")
plt.subplots_adjust(left=0.3)
```