

HW 3 Simple Linear Regression

Importing Google Drive Installing Process Mining for Python and Graphviz

```
[1] from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

!pip install pm4py
!apt install -y graphviz
!pip install graphviz

Collecting pm4py
  Downloading pm4py-2.7.0-py3-none-any.whl (1.8 MB)
    1.8/1.8 MB 39.9 MB/s eta 0:00:00
Collecting deprecation (from pm4py)
  Downloading deprecation-2.1.0-py2.py3-none-any.whl (11 kB)
Requirement already satisfied: graphviz in /usr/local/lib/python3.10/dist-packages (from pm4py) (0.20.1)
Collecting intervaltree (from pm4py)
  Downloading intervaltree-3.1.0.tar.gz (32 kB)
  Preparing metadata (setup.py) ... done
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from pm4py) (4.9.3)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-packages (from pm4py) (3.7.1)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from pm4py) (3.1)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from pm4py) (1.23.5)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from pm4py) (1.5.3)
Requirement already satisfied: pydotplus in /usr/local/lib/python3.10/dist-packages (from pm4py) (2.0.2)
Requirement already satisfied: pytz in /usr/local/lib/python3.10/dist-packages (from pm4py) (2023.3.post1)
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from pm4py) (1.11.2)
Collecting stringdist (from pm4py)
  Downloading StringDist-1.0.9.tar.gz (7.4 kB)
  Preparing metadata (setup.py) ... done
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from pm4py) (4.66.1)
Requirement already satisfied: cvxopt in /usr/local/lib/python3.10/dist-packages (from pm4py) (1.3.2)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from deprecation->pm4py) (23.1)
Requirement already satisfied: sortedcontainers<3.0,>=2.0 in /usr/local/lib/python3.10/dist-packages (from intervaltree->pm4py) (2.4.0)
Requirement already satisfied: contourpy==1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->pm4py) (1.1.0)
Requirement already satisfied: cycler==0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib->pm4py) (0.11.0)
Requirement already satisfied: fonttools==4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->pm4py) (4.42.1)
Requirement already satisfied: kiwisolver==1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->pm4py) (1.4.5)
Requirement already satisfied: pillow==6.2.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->pm4py) (9.4.0)
Requirement already satisfied: pyparsing==2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->pm4py) (3.1.1)
Requirement already satisfied: python-dateutil==2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib->pm4py) (2.8.2)
Requirement already satisfied: six==1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil->matplotlib->pm4py) (1.16.0)
Building wheels for collected packages: intervaltree, stringdist
  Building wheel for intervaltree (setup.py) ... done
  Created wheel for intervaltree: filename=intervaltree-3.1.0-py2.py3-none-any.whl size=26094 sha256=66e7f15ced28da38b2e9f60090189cd859a65ebcf7557da47453669ceedab52a
  Stored in directory: /root/.cache/pip/wheels/fa/80/8c/43488a924a046b733b64de3fac99252674c892a4c3801c0a61
  Building wheel for stringdist (setup.py) ... done
  Created wheel for stringdist: filename=StringDist-1.0.9-cp310-cp310-linux_x86_64.whl size=16968 sha256=4272309d6d7be07dfbb19bb18e6ead86653ab6a50f5f94709fa9ea76c18565
  Stored in directory: /root/.cache/pip/wheels/79/25/0a/4a0c2c7af31a8b473f6e10e30312390f31ae89972260ed7b99
Successfully built intervaltree stringdist
Installing collected packages: stringdist, intervaltree, deprecation, pm4py
Successfully installed deprecation-2.1.0 intervaltree-3.1.0 pm4py-2.7.0 stringdist-1.0.9
```

Utilizing Numpy and Matplot for graph
Pandas utilized for analysis on data
Importing the dataset
Splitting the dataset to training and test values
Linear Regression is then implemented to training values
Test results are then predicted using the y_predictor variable

```
[3] import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

[ ] #Importing dataset

[4] dataset = pd.read_csv('/content/drive/My Drive/Process mining project/Salary_Data.csv')
X_Axis = dataset.iloc[:, :-1].values
Y_Axis = dataset.iloc[:, -1].values

[ ] #Splitting Dataset into the training set and test set

[5] from sklearn.model_selection import train_test_split
X_training, X_test, y_training, y_test = train_test_split(X_Axis, Y_Axis, test_size = 1/3, random_state = 0)

[ ] #Training the Simple Linear Regression Model on the Training Set

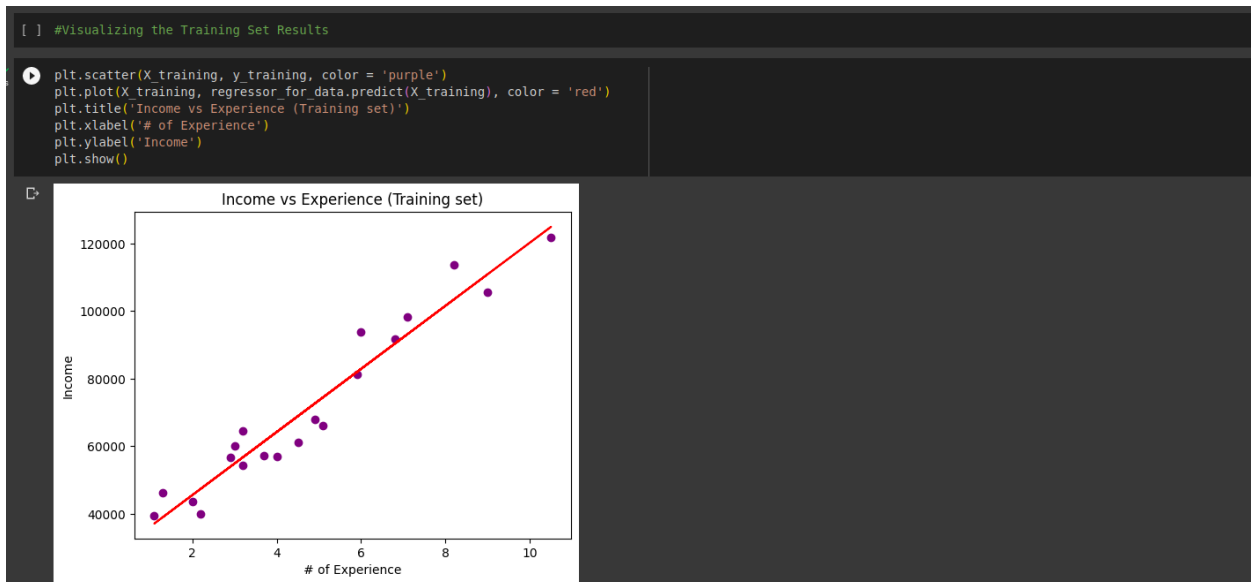
[6] from sklearn.linear_model import LinearRegression
regressor_for_data = LinearRegression()
regressor_for_data.fit(X_training, y_training)

* LinearRegression
LinearRegression()

[ ] #Predicting Test Results

[7] y_predictor = regressor_for_data.predict(X_test)
```

Values of training are now shown using the scatter, plot, title, xlabel, ylabel and show functions



Using the same functions we use test variables to visualize the dataset

