Ex. No.: 6 Write a program for Polynomial Regression Algorithm

AIM:

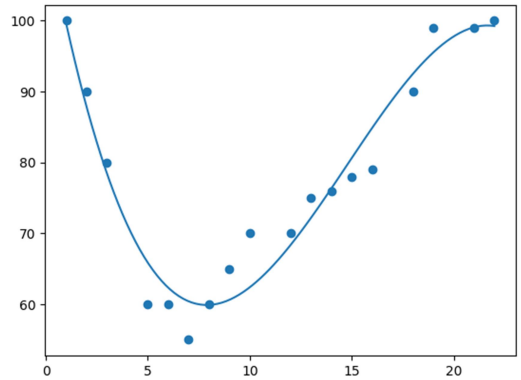
To write a Python program to calculate the Polynomial regression by using the given series and csv files.

Polynomial Regression: If data points clearly will not fit a linear regression (a straight line through all data points), it might be ideal for polynomial regression. Polynomial regression, like linear regression, uses the relationship between the two variables to find the best way to draw a line through the data points. **For example:** We have registered 18 cars as they were passing a certain tollbooth. We have registered the car's speed and the time of day (hour) the passing occurred. The x-axis represents the hours of the day and the y-axis represents the speed.

(i) Program (By using the given series):

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.metrics import r2_score
x = [1,2,3,5,6,7,8,9,10,12,13,14,15,16,18,19,21,22]
y = [100,90,80,60,60,55,60,65,70,70,75,76,78,79,90,99,99,100]
mymodel = np.poly1d(np.polyfit(x, y, 3))
myline = np.linspace(1, 22, 100)
plt.scatter(x, y)
plt.plot(myline, mymodel(myline))
plt.show()
print(r2_score(y, mymodel(x)))
speed = mymodel(17)
print(speed)
```

Output:



R-Squared value = 0.9432150416451027. The result 0.94 shows that there is a very good relationship and we can use polynomial regression in future predictions. The predicted speed of a car passing at 17th hour is 88.87331269697984.

(ii) Program (By using the given series):

```
import numpy as np
from sklearn.metrics import r2_score
x = [89, 43, 36, 36, 95, 10, 66, 34, 38, 20, 26, 29, 48, 64, 6, 5, 36, 66, 72, 40]
y = [21, 46, 3, 35, 67, 95, 53, 72, 58, 10, 26, 34, 90, 33, 38, 20, 56, 2, 47, 15]
mymodel = np.poly1d(np.polyfit(x, y, 3))
print(r2_score(y, mymodel(x)))
```

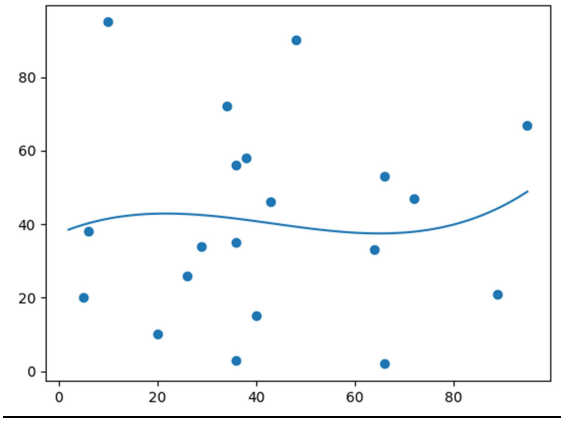
Output:

R-Squared value = 0.009952707566680652. The result 0.00995 indicates a very bad relationship and this data set is not suitable for polynomial regression.

(iii) Program (By using the given series):

```
import numpy as np
import matplotlib.pyplot as plt
x = [89, 43, 36, 36, 95, 10, 66, 34, 38, 20, 26, 29, 48, 64, 6, 5, 36, 66, 72, 40]
y = [21, 46, 3, 35, 67, 95, 53, 72, 58, 10, 26, 34, 90, 33, 38, 20, 56, 2, 47, 15]
mymodel = np.poly1d(np.polyfit(x, y, 3))
myline = np.linspace(2, 95, 100)
plt.scatter(x, y)
plt.plot(myline, mymodel(myline))
plt.show()
```

Output:

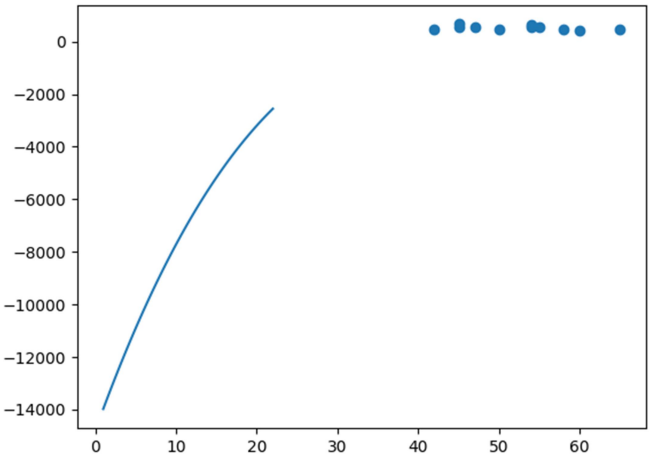


These values are very bad fit for polynomial regression and would not be the best method to predict future values.

(iv) Program (By using the given csv file):

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.metrics import r2_score
df=pd.read_csv("Book.csv", header=0, sep=",")
x=df["Duration"]
y=df["calories"]
mymodel = np.poly1d(np.polyfit(x, y, 3))
myline = np.linspace(1, 22, 100)
plt.scatter(x, y)
plt.plot(myline, mymodel(myline))
plt.show()
print(r2_score(y, mymodel(x)))
```

Output:



Here, R-Squared value is 0.40379738147486666. The values of duration and calories are not good to fit for polynomial regression and would not be the best method to predict future values.