

**Rubric of Lab2**

Name: <b>Joud Mohammed Alahmari</b> ID: <b>2008071</b>	<b>Max Mark s</b>	<b>Obtaine d Marks</b>	Comments
<b>CONTENTS: Question 1</b>			
<b>Question 1(1)</b>	2		
<b>Question 1(2)</b>	6		
<b>Question 1(3)</b>	3		
<b>Question 1(4)</b>	2		
<b>Question 1(5)</b>	2		
<b>Question 1(6)</b>	2		
<b>Question 1(7)</b>	2		
<b>TOTAL MARKS</b>	<b>19</b>		
<b>ORGANISATION</b>			
<ul style="list-style-type: none"> <li>PDF or DOCX file and source code file, Name your file as &lt;Assignment1_ID&gt;, cover page body, headings, fonts, style, exhibits, length, page numbers, etc.</li> </ul>	2		

<ul style="list-style-type: none"> <li>• Informative Screen shots, tables, figures, etc..</li> </ul>	2		
<b>TOTAL MARKS</b>	<b>23</b>		
<b>PENALTY</b>			
<ul style="list-style-type: none"> <li>• Similarity index</li> <li>• Late submission</li> </ul>			
<b>FINAL MARKS</b>	<b>15</b>		

### **Answer all questions:**

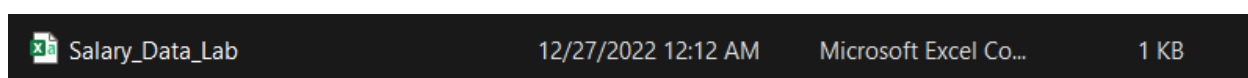
### **Document Requirements :**

- ❖ The Lab is submitted electronically through Blackboard. Each late assignment will be penalized 1 point per late working day.
- ❖ The format of document:
  - ✓ Softcopy of document as DOCX file and source code file. Name your file as < Lab5 \_ID>. send me it by using blackboard
  - ✓ Cover page with the members' names, ID
  - ✓ Text times roman font 12 or equivalent
  - ✓ Line spacing 1.5 and must include page numbering
  - ✓ Properly bound (Do not submit loose pages in folders. The instructor will not be responsible if any of the pages are missing).
  - ✓ Indent the first line of all paragraphs
  - ✓ Justify all paragraphs

## Question 1

You are working in a specific company as a data analyst. This company has an available position for country manager and new person applied for this position. He/She has been working for 3 years as a country manager in another company and his/her salary is 133,000\$. HR sent you the positions' salary dataset and ask you if he/she deserve the 133,000\$? Consider his/her level on 5.5.

1. Download the attached dataset.



2. Predict the new country manager's salary using the following:

- a. Simple linear regression

```
Spyder (Python 3.9)
File Edit Search Source Run Debug Consoles Projects Tools View Help

C:\Users\joudm\Desktop\ML\lab2_part1_2008071.py

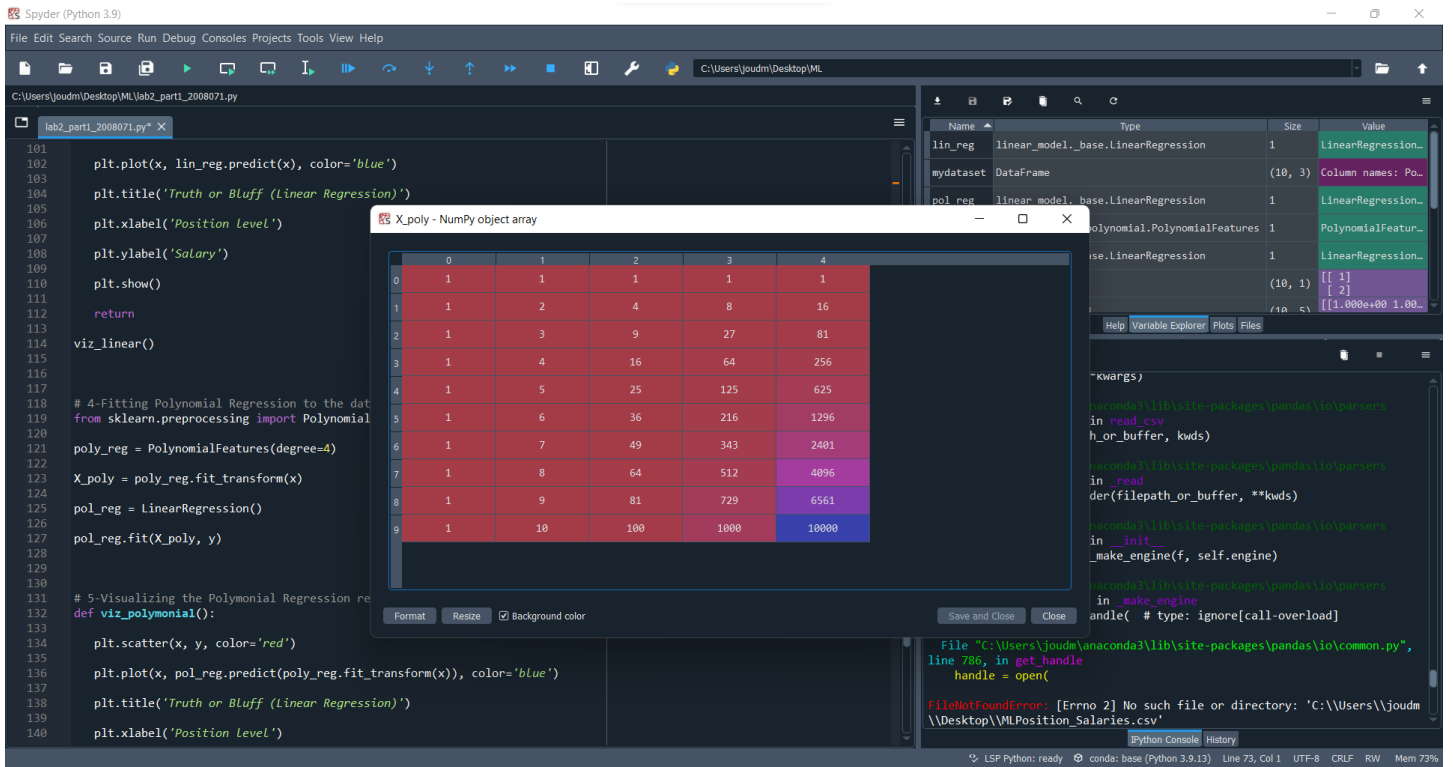
lab2_part1_2008071.py
1  #-*- coding: utf-8 -*-
2  """
3  Created on Tue Dec 27 00:15:17 2022
4
5  @author: Joud
6  """
7  # 1-import libraries
8  import numpy as np
9  import matplotlib.pyplot as plt
10 import pandas as pd
11
12 # 2-import dataSet
13 mydataset = pd.read_csv("Salary_Data_Lab.csv")
14 mydataset= pd.read_csv("C:\\Users\\joudm\\Desktop\\ML\\Salary_Data_Lab.csv")
15
16 # 3-select varib
17 x = mydataset.iloc[:, :-1].values
18 y = mydataset.iloc[:, 1].values
19
20 # 4-splitting the dataset
21 from sklearn.model_selection import train_test_split
22 X_train, X_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 0 )
23
24
25 # 5-training the module on training set
26 from sklearn.linear_model import LinearRegression
27 regressor = LinearRegression()
28 regressor.fit(X_train, y_train)
29
30 # 6-predicting the test set result
31 y_pred = regressor.predict(X_test)
32
33
34
35
36
37
38
39
40
```

Name	Type	Size	Value
lin_reg	linear_model_base.LinearRegression	1	LinearRegress...
mydataset	DataFrame	(10, 3)	Column names:..
pol_reg	linear_model_base.LinearRegression	1	LinearRegress...
poly_reg	preprocessing_polynomial.PolynomialFeatures	1	PolynomialFea...
regressor	linear_model_base.LinearRegression	1	LinearRegress...
x	Array of int64	(10, 1)	[[ 1]
X_poly	Array of float64	(10, 5)	[[1.000e+00 1.
X_test	Array of float64	(9, 1)	[[10.7]
X_train	Array of float64	(35, 1)	[[11. ]

```
Console 1/A X
runfile('C:/Users/joudm/Desktop/ML/Lab2_part1(elan).py',
self.handle = get_handle( # type: ignore[call-overload]
File "C:/Users/joudm/anaconda3/lib/site-packages/pandas/io
common.py", line 786, in get_handle
handle = open(
FileNotFoundError: [Errno 2] No such file or directory: 'C:/Users/
joudm/Desktop/MLPosition_Salaries.csv'
In [11]: runfile('C:/Users/joudm/Desktop/ML/Lab2_part1(elan).py',
wdir='C:/Users/joudm/Desktop/ML')
In [12]: runfile('C:/Users/joudm/Desktop/ML/Lab2_part1_2008071.py',
wdir='C:/Users/joudm/Desktop/ML')
In [13]:
```

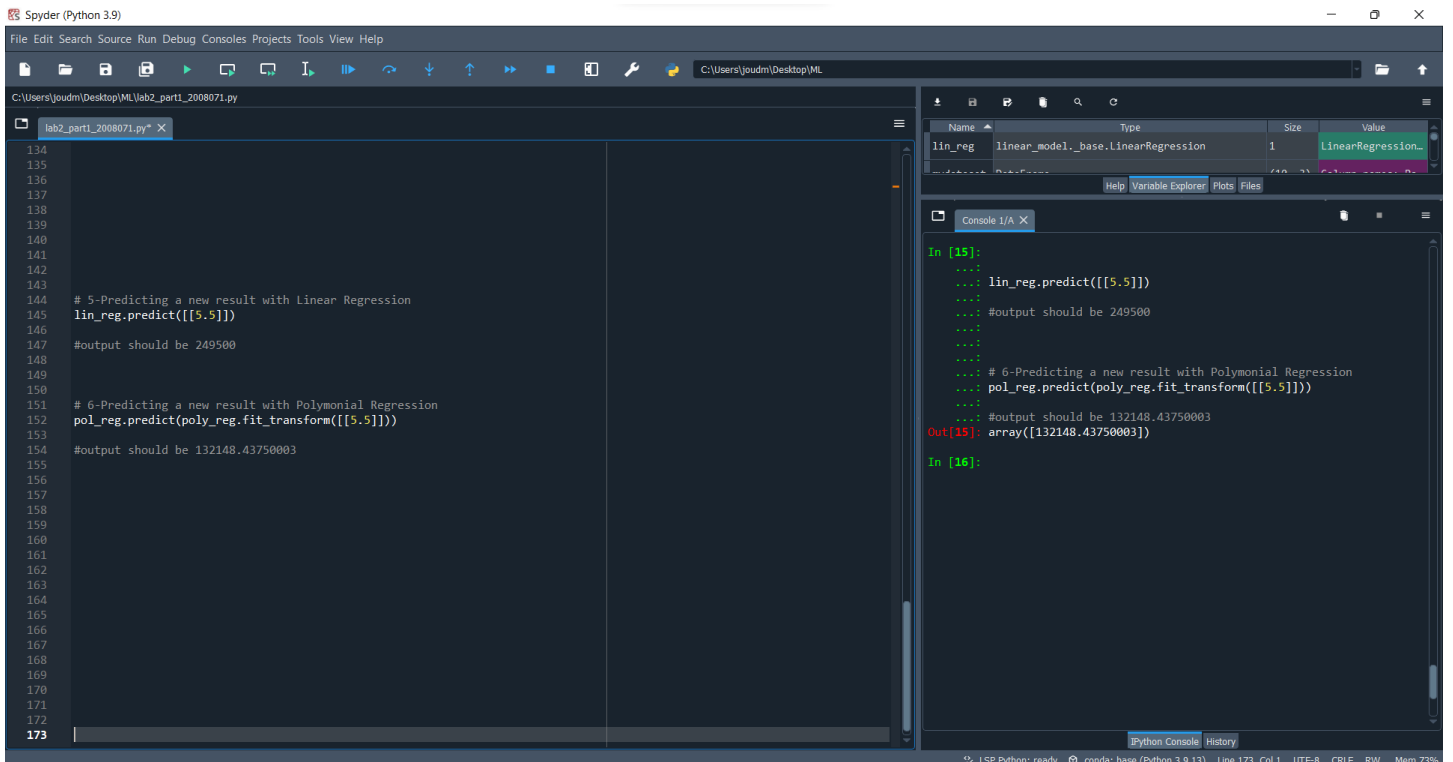
LSP Python: ready conda: base (Python 3.9.13) Line 38, Col 1 UTF-8 CRLF RW Mem 73%





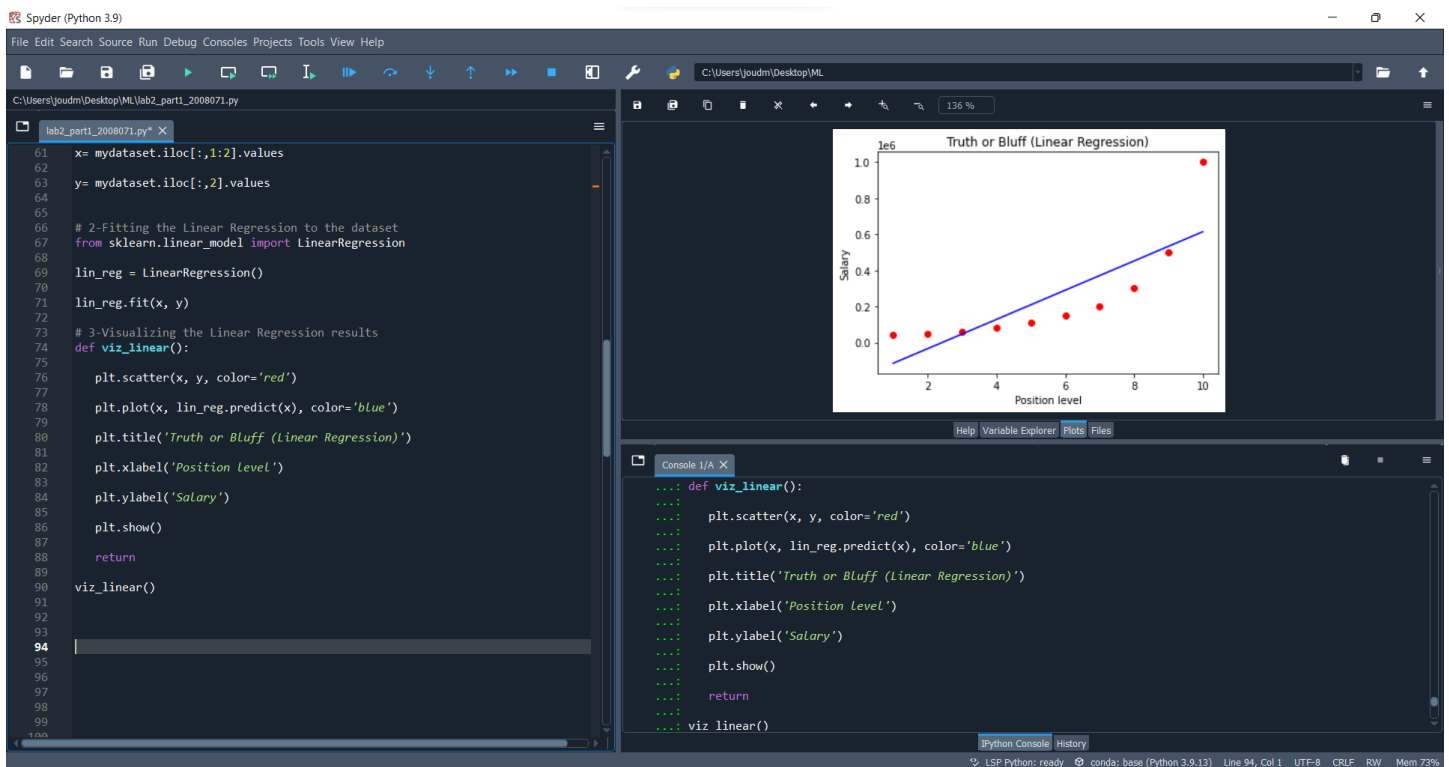
c. Support vector regression.

3. Write or show screenshot of the prediction results.

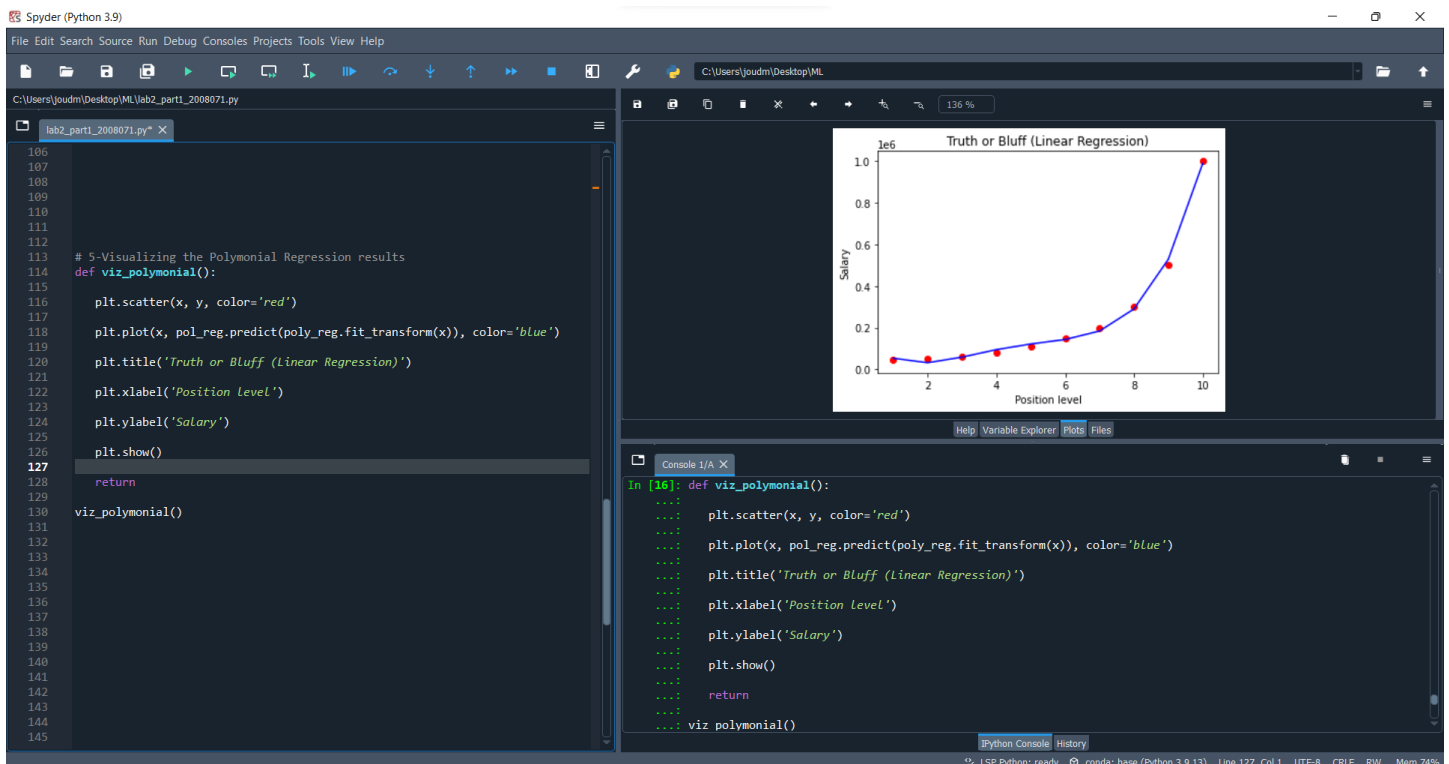


4. Visualize the simple linear regression result (show screenshot).

## 5. Visualize the polynomial linear regression results (show screenshot).



## 6. Visualize the support vector regression results (show screenshot).



# Support Vector Regression (SVR)

The screenshot shows a Jupyter Notebook interface with a code editor on the left and variable explorer/output areas on the right.

**Code Editor (untitled3.py):**

```
1  #-*- coding: utf-8 -*-
2  """
3  Created on Mon Jan  2 15:45:21 2023
4
5  @author: joudm
6  """
7
8  # 1-import Libraries
9  import numpy as np
10 import matplotlib.pyplot as plt
11 import pandas as pd
12
13 # 2-import dataSet
14 mydataset= pd.read_csv("C:\\Users\\joudm\\Desktop\\ML\\Position_Salaries.csv")
15 X = mydataset.iloc[:, 1:2].values
16 y = mydataset.iloc[:, 2].values
17
18 from sklearn.preprocessing import StandardScaler
19 sc_X = StandardScaler()
20 sc_y = StandardScaler()
21 X = sc_X.fit_transform(X)
22 y = np.ravel(sc_y.fit_transform(y.reshape(-1,1)))#considerit as an array and res...
```

**Variable Explorer:**

Name	Type	Size	Value
mydataset	DataFrame	(10, 3)	Column names: Position, Level, Salary
preprocessing_data.StandardScaler	StandardScaler	1	StandardScaler object of...
preprocessing_data.StandardScaler	StandardScaler	1	StandardScaler object of...
Array of float64	Array of float64	(10, 1)	[[-1.5666989 ] [-1.21854359]]
		(10,)	[-0.72804253 -0.70243757...]

**NumPy object array windows:**

- X - NumPy object array:** A 10x1 matrix of standardized features.
- y - NumPy object array:** A 10x1 matrix of standardized target values.

**Console Output:**

```
In [34]: runfile('C:/Users/joudm/Desktop/ML/untitled3.py', wdir='C:/Users/joudm/Desktop/ML')
```

Spyder (Python 3.9)

File Edit Search Source Run Debug Consoles Projects Tools View Help

C:\Users\joudm\Desktop\ML

C:\Users\joudm\Desktop\ML\lab2\_SVR\_2008071\_ML.py

```

1  """
2  Created on Mon Jan 2 15:45:21 2023
3
4  @author: joudm
5  """
6
7
8  # 1-import Libraries
9  import numpy as np
10 import matplotlib.pyplot as plt
11 import pandas as pd
12
13 # 2-import dataSet
14 mydataset= pd.read_csv("C:\\Users\\joudm\\Desktop\\ML\\Position_Salari
15 X = mydataset.iloc[:, 1:2].values
16 y = mydataset.iloc[:, 2].values
17
18 from sklearn.preprocessing import StandardScaler
19 sc_X = StandardScaler()
20 sc_y = StandardScaler()
21 X = sc_X.fit_transform(X)
22 y = np.ravel(sc_y.fit_transform(y.reshape(-1,1)))#considerit as an arr
23
24 from sklearn.svm import SVR #it will add any kind of help
25 regressor = SVR(kernel = 'rbf')
26 regressor.fit(X,y)
27
28 y_pred= sc_y.inverse_transform(regressor.predict(sc_X.transform([[6.5]])).reshape(-1,1)) #we will predict the value of y /we
29
30 plt.scatter(X,y,color='red')
31 plt.plot(X, regressor.predict(X),color = 'blue')
32 plt.title('Truth or Bluff (SVR)')
33 plt.xlabel('Position Level')
34 plt.ylabel('Salary')
35 plt.show()
36
37
38
39
40

```

lab2\_SVR\_2008071\_ML.py X

0

170370

Format Resize Background color Save and Close Close

Name	Type	Size	Value
mydataset	DataFrame	(10, 3)	Column names: Position, Level, Salary
regressor	svm._classes.SVR	1	SVR object of sklearn.sv-
sc_X	preprocessing_data.StandardScaler	1	StandardScaler object of-
sc_y	preprocessing_data.StandardScaler	1	StandardScaler object of-
X	Array of float64	(10, 1)	[[-1.5666989 ] [-1.21854359]
y	Array of float64	(10, 1)	[-0.72004253 -0.70243757-
y_pred	Array of float64	(1, 1)	[[170370.0204065]]

Help Variable Explorer Plots Files

Console I/A X

File "C:\Users\joudm\anaconda3\lib\site-packages\sklearn\utils\validation.py", line 746, in check\_array  
X = check\_array(  
File "C:\Users\joudm\anaconda3\lib\site-packages\sklearn\utils\validation.py", line 746, in check\_array  
array = np.asarray(array, order=order, dtype=dtype)  
TypeError: float() argument must be a string or a number, not 'builtin\_function\_or\_method'  
  
In [37]: runfile('C:/Users/joudm/Desktop/ML/untitled3.py', wdir='C:/Users/joudm/Desktop/ML')  
  
In [38]:

Python Console History

LSP Python: ready conda: base (Python 3.9.13) Line 28, Col 154 UTF-8 CRLF RW Mem 58%

Spyder (Python 3.9)

File Edit Search Source Run Debug Consoles Projects Tools View Help

C:\Users\joudm\Desktop\ML

C:\Users\joudm\Desktop\ML\lab2\_SVR\_2008071\_ML.py

```

2  """
3  Created on Mon Jan 2 15:45:21 2023
4
5  @author: joudm
6  """
7
8  # 1-import Libraries
9  import numpy as np
10 import matplotlib.pyplot as plt
11 import pandas as pd
12
13 # 2-import dataSet
14 mydataset= pd.read_csv("C:\\Users\\joudm\\Desktop\\ML\\Position_Salaries.csv")
15 X = mydataset.iloc[:, 1:2].values
16 y = mydataset.iloc[:, 2].values
17
18 from sklearn.preprocessing import StandardScaler
19 sc_X = StandardScaler()
20 sc_y = StandardScaler()
21 X = sc_X.fit_transform(X)
22 y = np.ravel(sc_y.fit_transform(y.reshape(-1,1)))#considerit as an array and reshape the y as 2D vector
23
24 from sklearn.svm import SVR #it will add any kind of help
25 regressor = SVR(kernel = 'rbf')
26 regressor.fit(X,y)
27
28 y_pred= sc_y.inverse_transform(regressor.predict(sc_X.transform([[6.5]])).reshape(-1,1)) #we will predict the value of y /we
29
30 plt.scatter(X,y,color='red')
31 plt.plot(X, regressor.predict(X),color = 'blue')
32 plt.title('Truth or Bluff (SVR)')
33 plt.xlabel('Position Level')
34 plt.ylabel('Salary')
35 plt.show()
36
37
38
39
40

```

lab2\_SVR\_2008071\_ML.py X

Truth or Bluff (SVR)

Help Variable Explorer Plots Files

Console I/A X

File "C:\Users\joudm\anaconda3\lib\site-packages\sklearn\utils\validation.py", line 746, in check\_array  
X = check\_array(  
File "C:\Users\joudm\anaconda3\lib\site-packages\sklearn\utils\validation.py", line 746, in check\_array  
array = np.asarray(array, order=order, dtype=dtype)  
TypeError: float() argument must be a string or a number, not 'builtin\_function\_or\_method'  
  
In [37]: runfile('C:/Users/joudm/Desktop/ML/untitled3.py', wdir='C:/Users/joudm/Desktop/ML')  
  
In [38]:

Python Console History

LSP Python: ready conda: base (Python 3.9.13) Line 28, Col 154 UTF-8 CRLF RW Mem 61%



7. Based on the scenario question, write your opinion about every model's result.

**A linear regression line does not easily fit the first graph. In order to achieve a minimum value of error and obtain the best estimate of salary based on the employee's entire position level, we used polynomial regression to fit polynomial line at the second graph (5.5)**

**The Polynomial linear regression module provides the best result of salary, which is as the two modules demonstrate, where ( 132148.43750003 ).**

**We can determine that the new employee deserves \$133,000 based on the module's outcome.**

**And I have notice that the SVR is sensitive to outlairs, it does not care about the data outside the margin,it try to find the line that keeps as many points as possible within epsilon. But outliers will still affect.**

**Good Luck ☺**