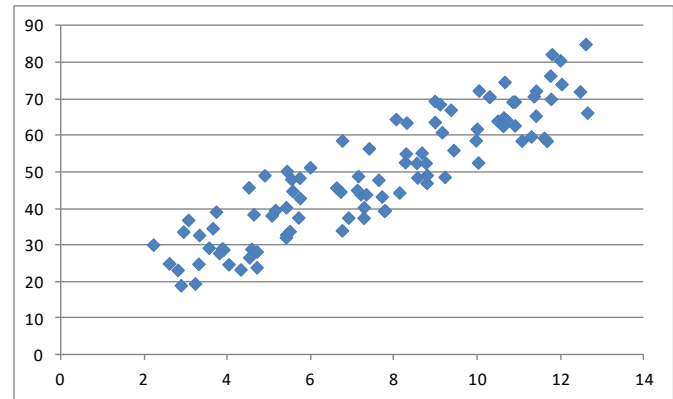
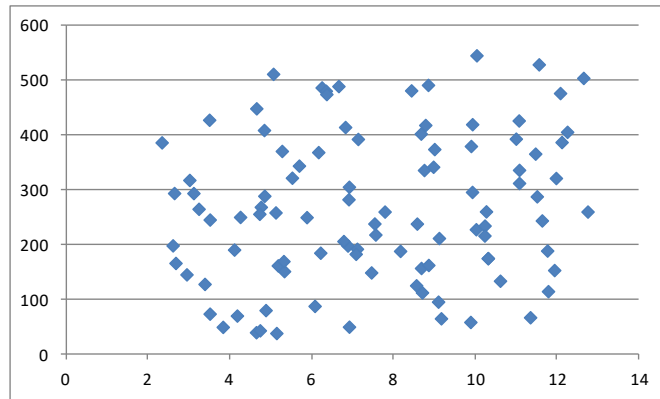


Diplomatura en Python aplicado a la ciencia de datos

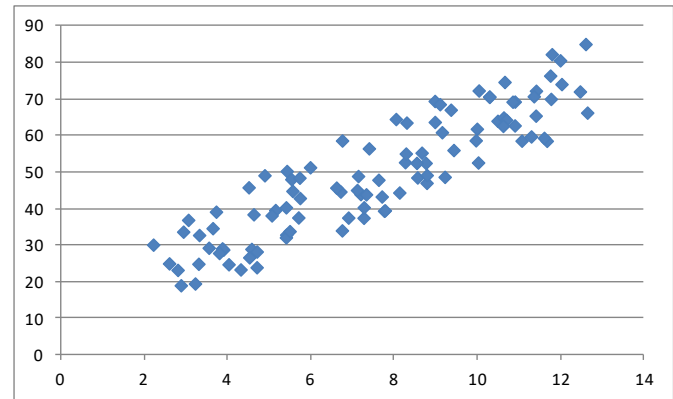
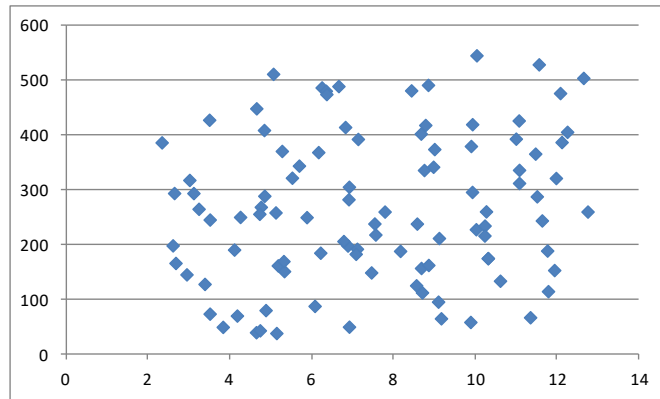
Correlaciones y Regresiones



Correlaciones



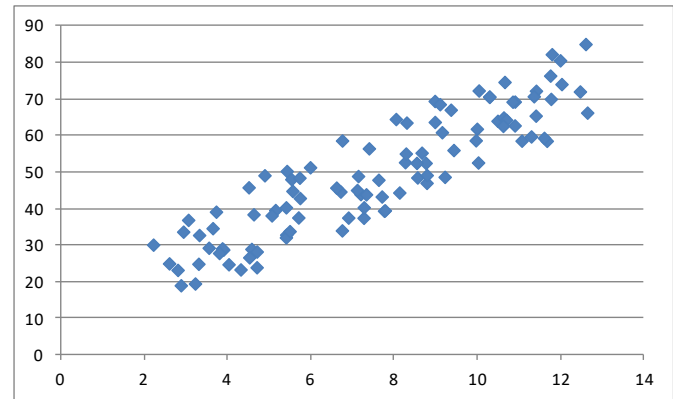
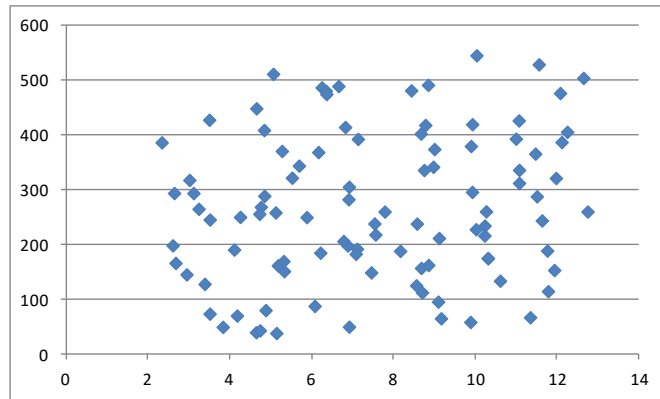
Correlaciones



$$C = \frac{\sum_{i=1}^{100} (x_i - x_{medio}) * (y_i - y_{medio})}{\sqrt{\sum_{i=1}^{100} (x_i - x_{medio})^2} * \sqrt{\sum_{i=1}^{100} (y_i - y_{medio})^2}}$$



Correlaciones



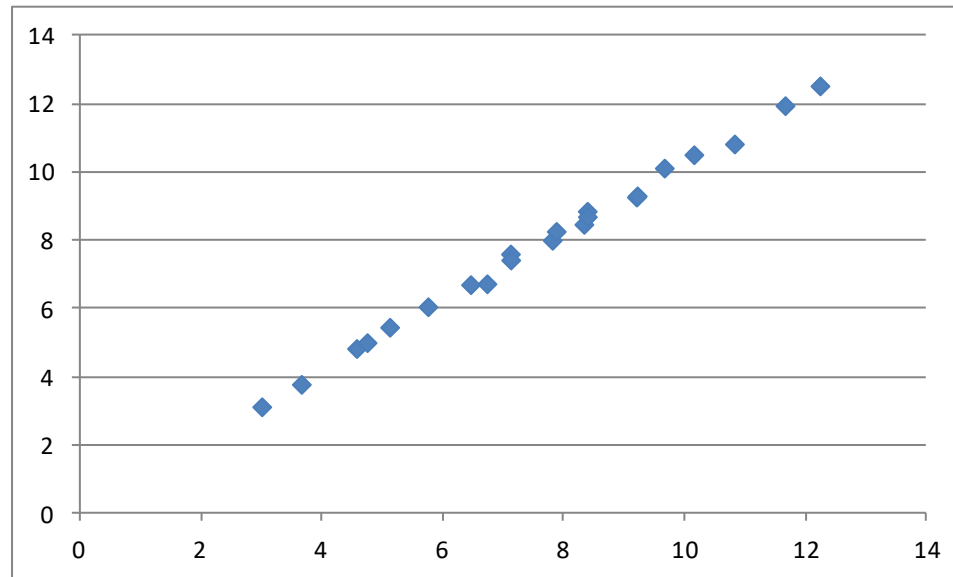
$$C = \frac{\sum_{i=1}^{100} (x_i - x_{medio}) * (y_i - y_{medio})}{\sqrt{\sum_{i=1}^{100} (x_i - x_{medio})^2} * \sqrt{\sum_{i=1}^{100} (y_i - y_{medio})^2}}$$

0.13

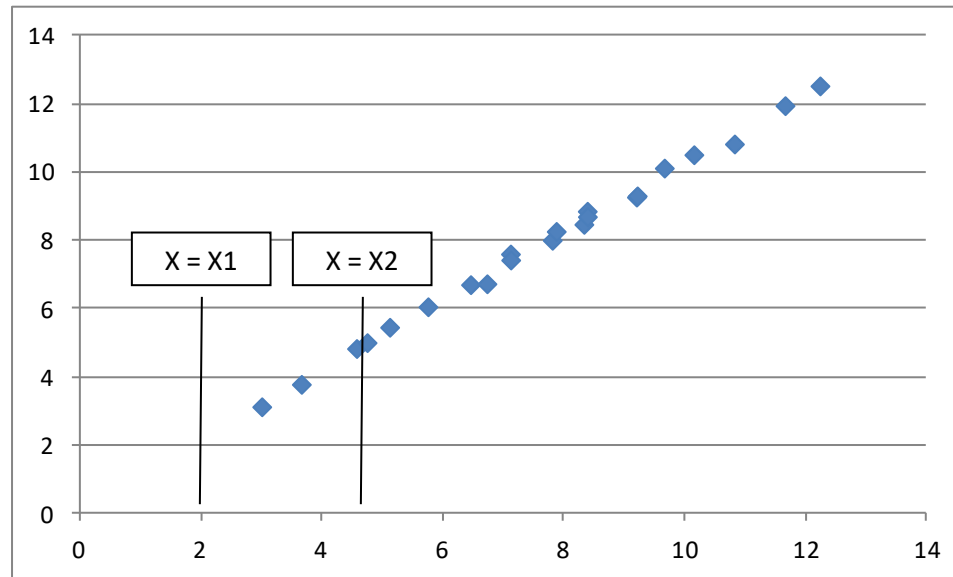
0.91



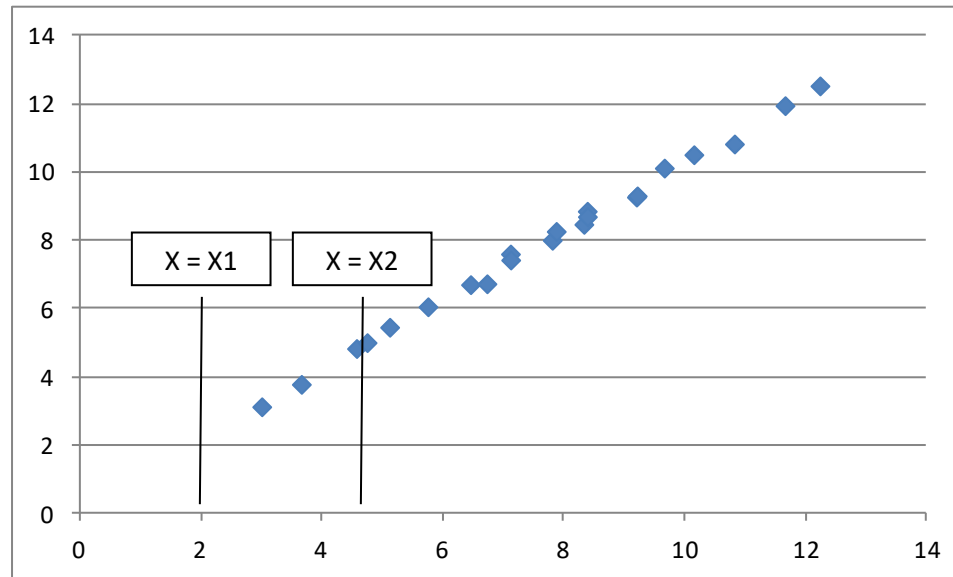
Regresiones



Regresiones



Regresiones



Tenemos: $\{(X_i; Y_i) \text{ tal que } 1 \leq i \leq N\}$

Queremos: $Y = a X + b$



Regresión Lineal Simple

$$\chi^2 = \sum_{i=1}^N (Y_i - aX_i - b)^2$$



Regresión Lineal Simple

$$\chi^2 = \sum_{i=1}^N (Y_i - aX_i - b)^2$$

$$\left\{ \begin{array}{l} \frac{\partial \chi^2}{\partial a} = 0 \\ \frac{\partial \chi^2}{\partial b} = 0 \end{array} \right\}$$



Regresión Lineal Simple

$$\chi^2 = \sum_{i=1}^N (Y_i - aX_i - b)^2$$

$$\left\{ \begin{array}{l} \frac{\partial \chi^2}{\partial a} = 0 \\ \frac{\partial \chi^2}{\partial b} = 0 \end{array} \right\}$$

$$\sigma_{jl} = \sum_{i=1}^N X_i^j Y_i^l$$



Regresión Lineal Simple

$$\chi^2 = \sum_{i=1}^N (Y_i - aX_i - b)^2$$

$$\left\{ \begin{array}{l} \frac{\partial \chi^2}{\partial a} = 0 \\ \frac{\partial \chi^2}{\partial b} = 0 \end{array} \right\}$$

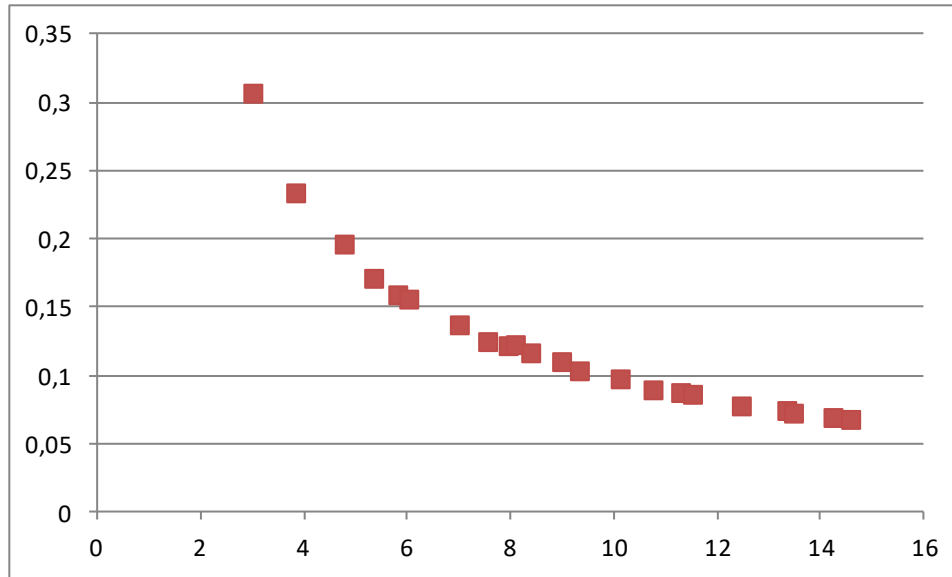
$$\sigma_{jl} = \sum_{i=1}^N X_i^j Y_i^l$$

$$a = \frac{\sigma_{10}\sigma_{01} - N\sigma_{11}}{\sigma_{10}^2 - N\sigma_{20}}$$

$$b = \frac{\sigma_{01} - a\sigma_{10}}{N}$$



Regresiones Linealizadas



$$Z_i = \frac{1}{Y_i}$$



Regresiones Polinómicas

$$Y = aX^2 + bX + c$$

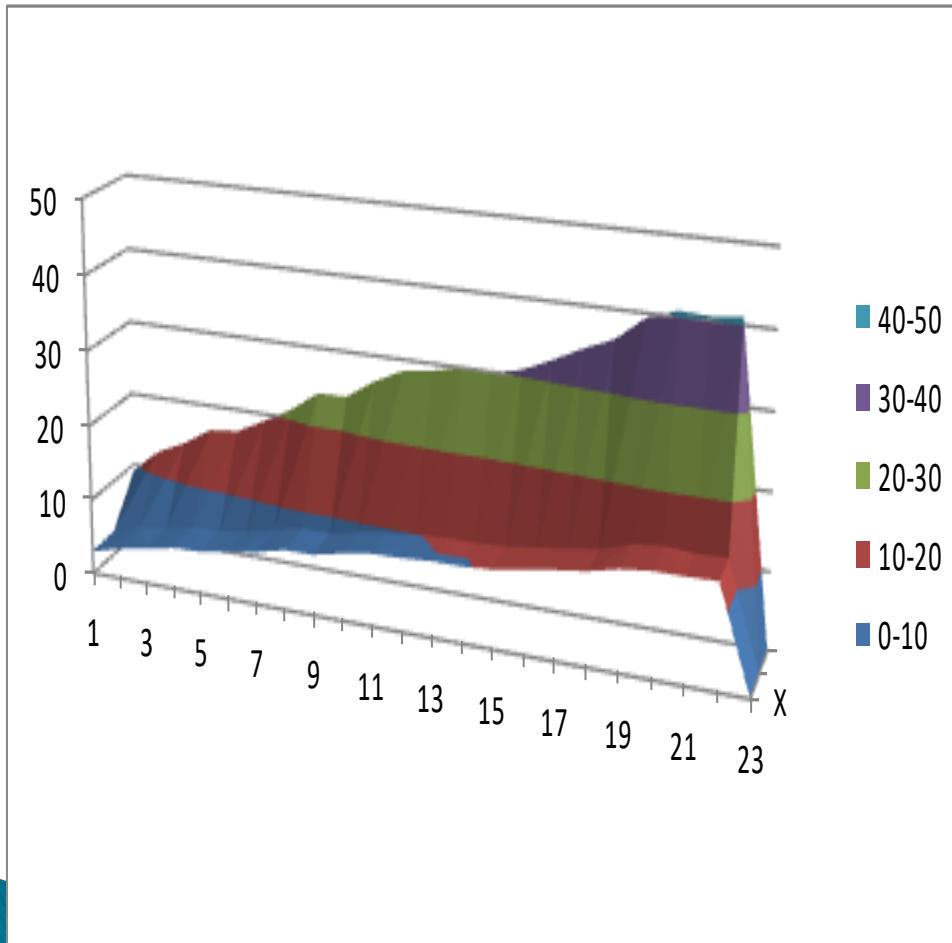
$$\chi^2 = \sum_{i=1}^N (Y_i - (aX_i^2 + bX_i + c))^2$$

$$\left\{ \begin{array}{l} \frac{\partial \chi^2}{\partial a} = 0 \\ \frac{\partial \chi^2}{\partial b} = 0 \\ \frac{\partial \chi^2}{\partial c} = 0 \end{array} \right\}$$

$$\begin{pmatrix} \sigma_{00} & \sigma_{10} & \sigma_{20} \\ \sigma_{10} & \sigma_{20} & \sigma_{30} \\ \sigma_{20} & \sigma_{30} & \sigma_{40} \end{pmatrix} x \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \sigma_{01} \\ \sigma_{11} \\ \sigma_{21} \end{pmatrix}$$



Regresión Multilineal



$$Z = aX + bY + c$$

$$\chi^2 = \sum_{i=1}^N (Z_i - (aX_i + bY_i + c))^2$$

$$\left\{ \begin{array}{l} \frac{\partial \chi^2}{\partial a} = 0 \\ \frac{\partial \chi^2}{\partial b} = 0 \\ \frac{\partial \chi^2}{\partial c} = 0 \end{array} \right\}$$

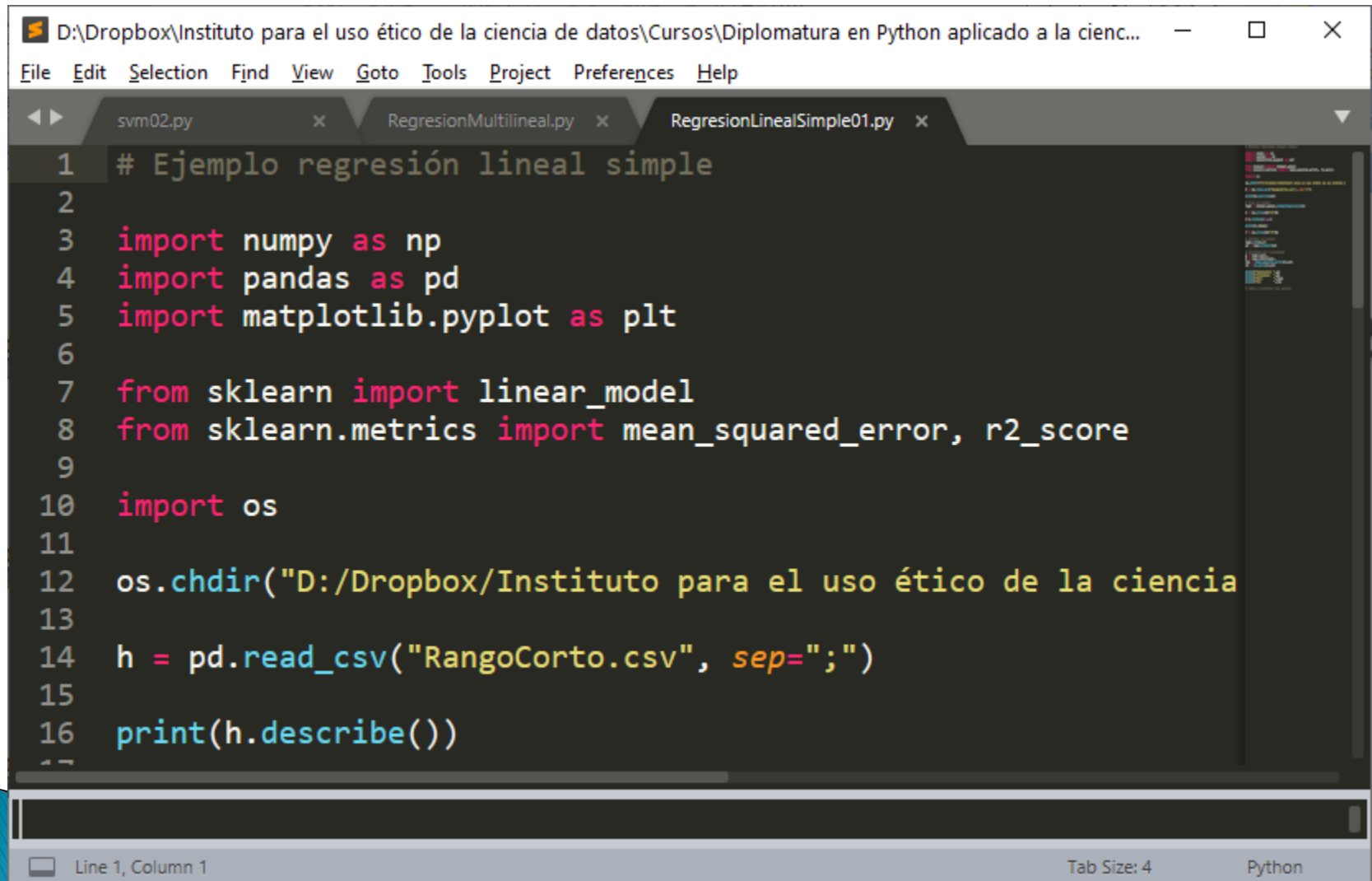


Implementación en Python

- ▶ Utilizamos los paquetes específicos:
 - Sklearn
 - Sklearn.metrics
- ▶ Y, por supuesto nos apoyamos en:
 - Numpy
 - Pandas
 - Matplotlib



Ejemplo de regresión lineal



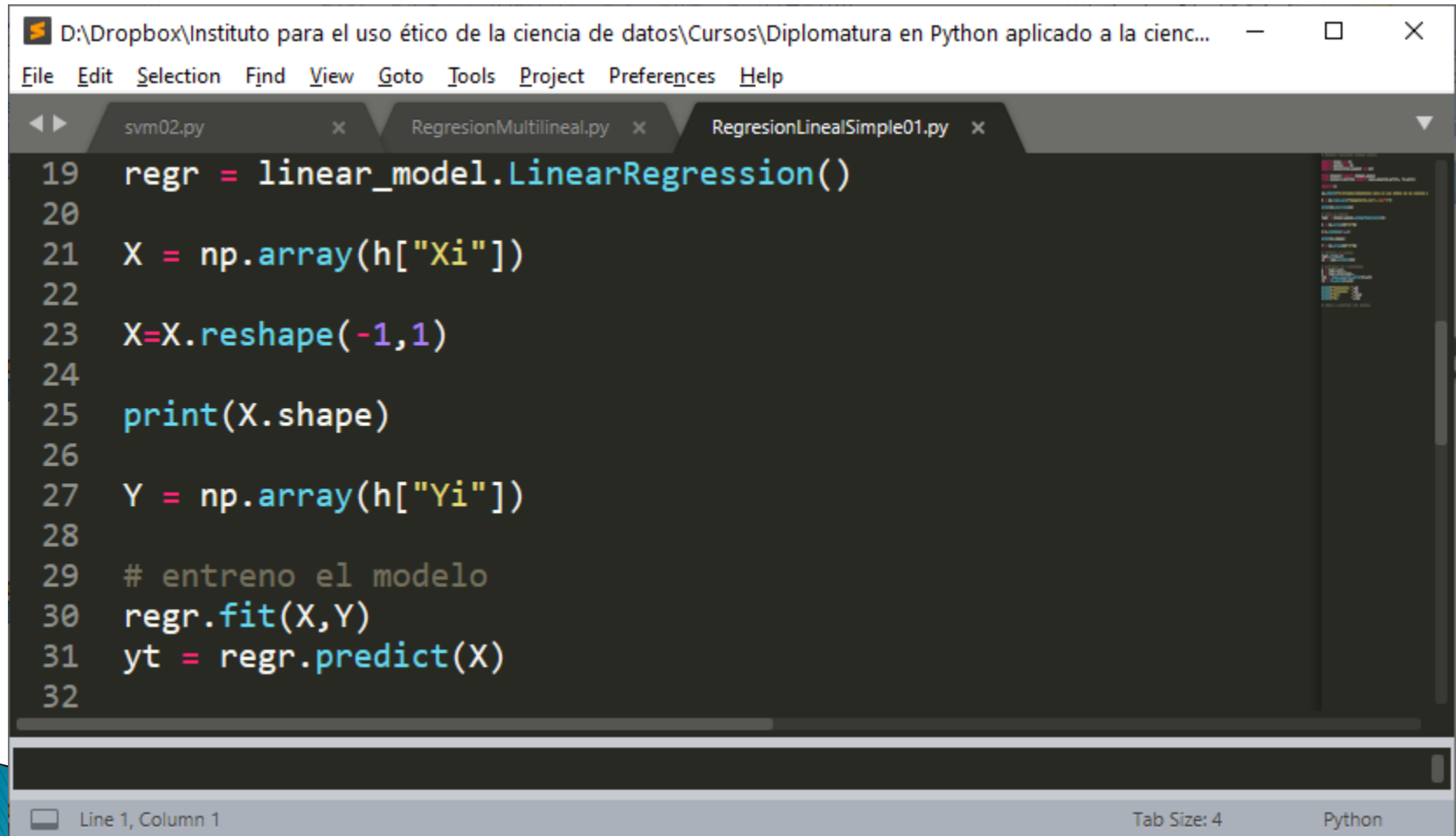
The image shows a screenshot of a code editor window. The title bar indicates the file path: D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia... The editor has three tabs open: svm02.py, RegresionMultilineal.py, and RegresionLinealSimple01.py. The active tab is RegresionLinealSimple01.py, which contains the following Python code:

```
1 # Ejemplo regresión lineal simple
2
3 import numpy as np
4 import pandas as pd
5 import matplotlib.pyplot as plt
6
7 from sklearn import linear_model
8 from sklearn.metrics import mean_squared_error, r2_score
9
10 import os
11
12 os.chdir("D:/Dropbox/Instituto para el uso ético de la ciencia
13
14 h = pd.read_csv("RangoCorto.csv", sep=";")
15
16 print(h.describe())
17
```

The status bar at the bottom shows "Line 1, Column 1", "Tab Size: 4", and "Python".



Ejemplo de regresión lineal



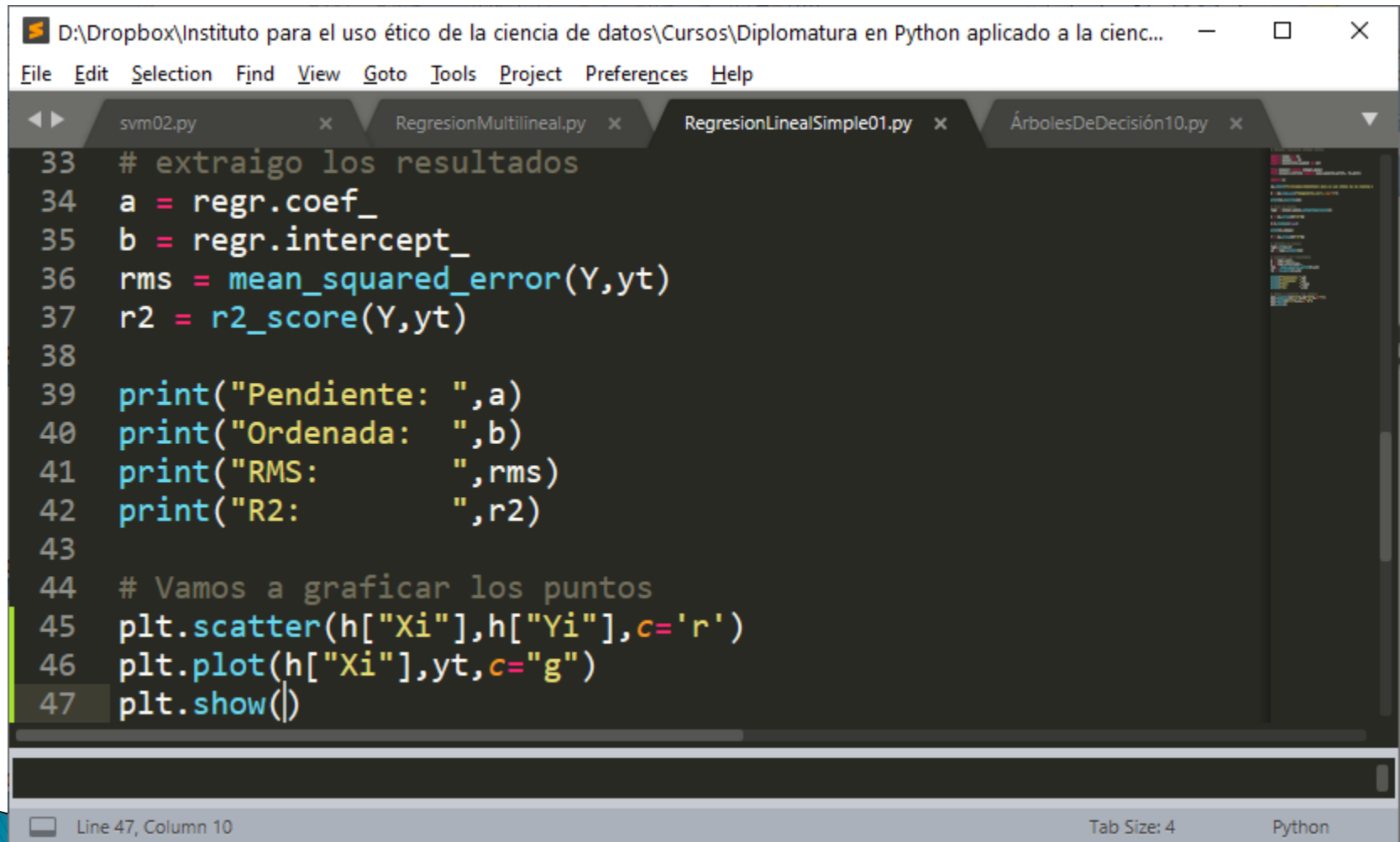
The image shows a screenshot of a Python IDE window. The title bar indicates the file path: D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia... The menu bar includes File, Edit, Selection, Find, View, Goto, Tools, Project, Preferences, and Help. The tab bar shows three open files: svm02.py, RegresionMultilineal.py, and RegresionLinealSimple01.py. The main editor area displays the following Python code:

```
19 regr = linear_model.LinearRegression()
20
21 X = np.array(h["Xi"])
22
23 X=X.reshape(-1,1)
24
25 print(X.shape)
26
27 Y = np.array(h["Yi"])
28
29 # entreno el modelo
30 regr.fit(X,Y)
31 yt = regr.predict(X)
32
```

The status bar at the bottom shows "Line 1, Column 1", "Tab Size: 4", and "Python".



Ejemplo de regresión lineal



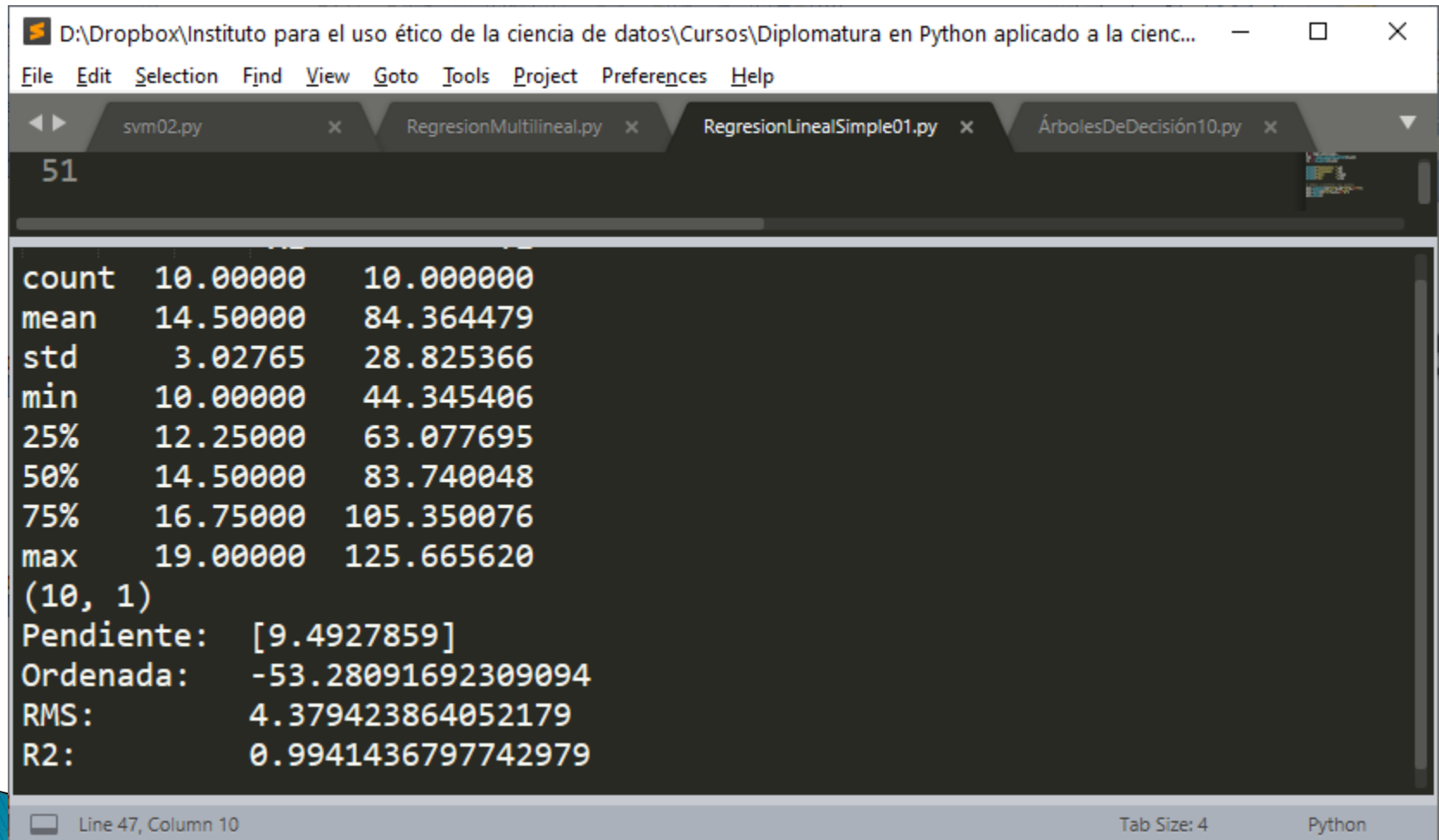
The image shows a screenshot of a Python IDE window. The title bar indicates the file path: D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia... The menu bar includes File, Edit, Selection, Find, View, Goto, Tools, Project, Preferences, and Help. The tab bar shows four open files: svm02.py, RegresionMultilineal.py, RegresionLinealSimple01.py, and ÁrbolesDeDecisión10.py. The main editor area displays Python code for linear regression. The code calculates the coefficients (a and b), the root mean square (rms), and the coefficient of determination (r2). It also prints these values and plots the data points (h["Xi"], h["Yi"]) as red dots, with the regression line (yt) as a green line. The status bar at the bottom shows "Line 47, Column 10", "Tab Size: 4", and "Python".

```
33 # extraigo los resultados
34 a = regr.coef_
35 b = regr.intercept_
36 rms = mean_squared_error(Y,yt)
37 r2 = r2_score(Y,yt)
38
39 print("Pendiente: ",a)
40 print("Ordenada: ",b)
41 print("RMS:      ",rms)
42 print("R2:       ",r2)
43
44 # Vamos a graficar los puntos
45 plt.scatter(h["Xi"],h["Yi"],c='r')
46 plt.plot(h["Xi"],yt,c="g")
47 plt.show()
```

Line 47, Column 10 Tab Size: 4 Python



Ejemplo de regresión lineal



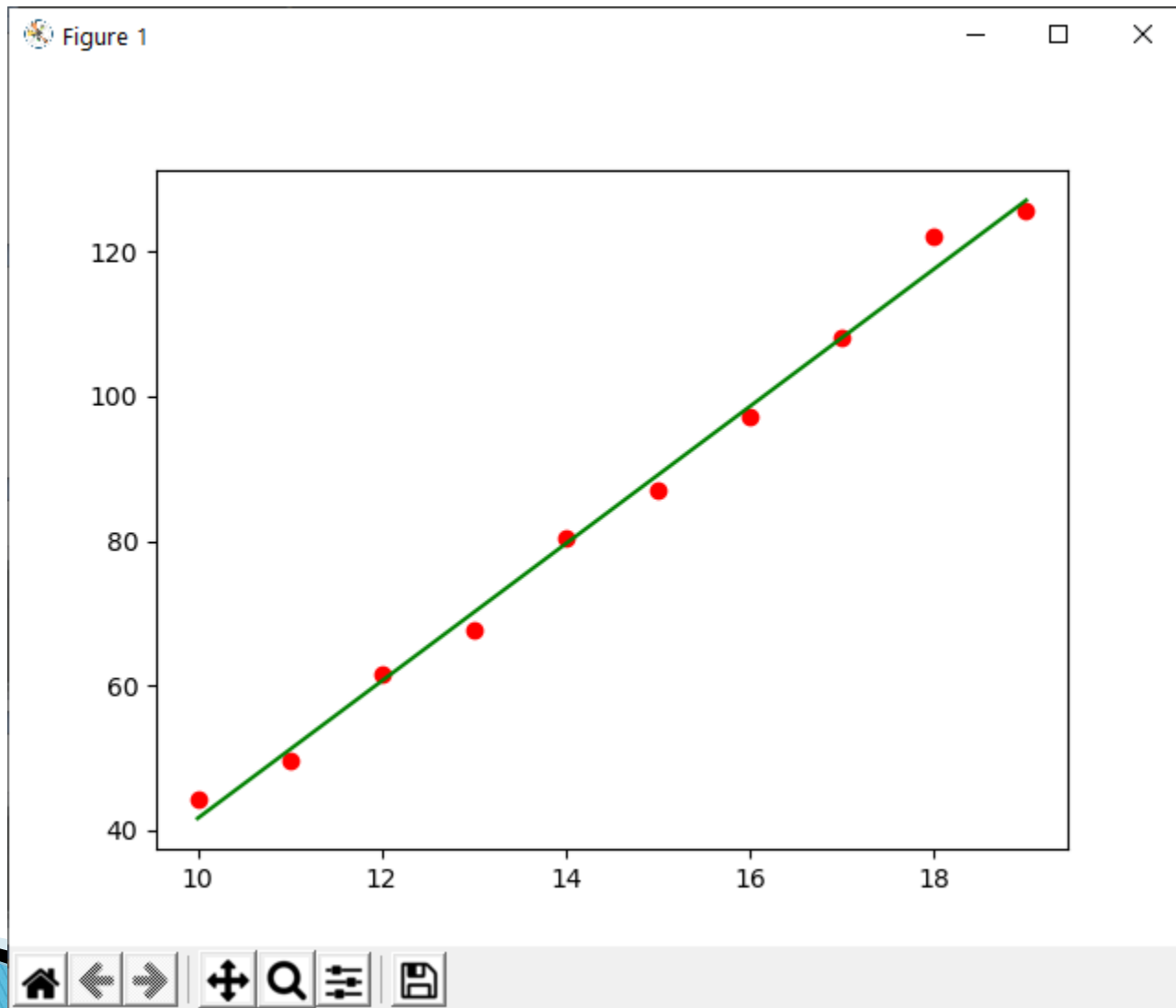
The screenshot shows a Python IDE window with the title bar "D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia...". The menu bar includes File, Edit, Selection, Find, View, Goto, Tools, Project, Preferences, and Help. The tab bar shows four open files: svm02.py, RegresionMultilineal.py, RegresionLinealSimple01.py, and ÁrbolesDeDecisión10.py. The main editor area displays the output of a linear regression model, starting with the number 51. The output includes a summary of the data and the regression coefficients.

```
51
count    10.00000    10.000000
mean     14.50000    84.364479
std       3.02765    28.825366
min      10.00000    44.345406
25%      12.25000    63.077695
50%      14.50000    83.740048
75%      16.75000   105.350076
max      19.00000   125.665620
(10, 1)
Pendiente: [9.4927859]
Ordenada:  -53.28091692309094
RMS:       4.379423864052179
R2:        0.9941436797742979
```

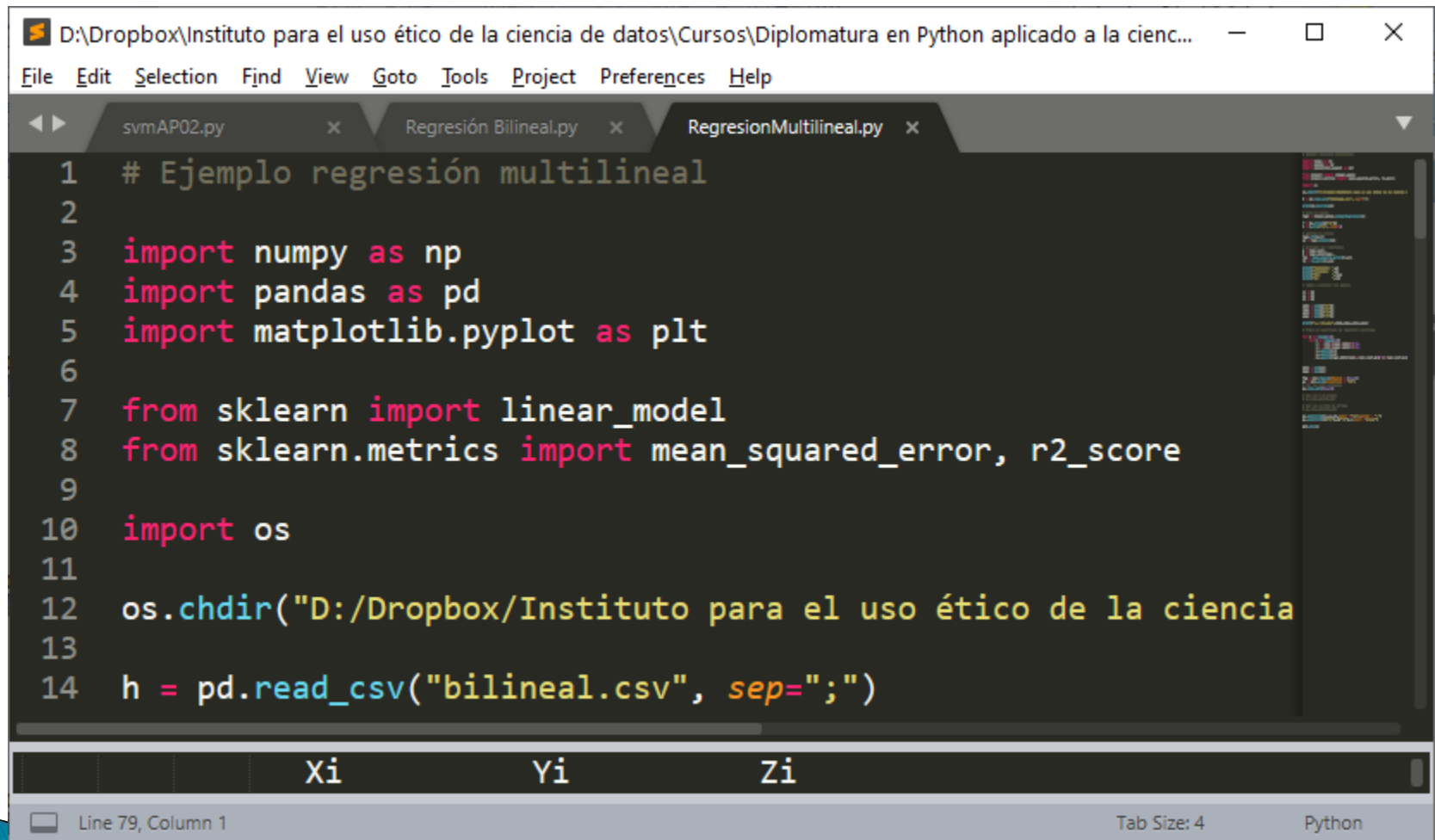
The status bar at the bottom indicates "Line 47, Column 10", "Tab Size: 4", and "Python".



Ejemplo de regresión lineal



Ejemplo de regresión multilineal



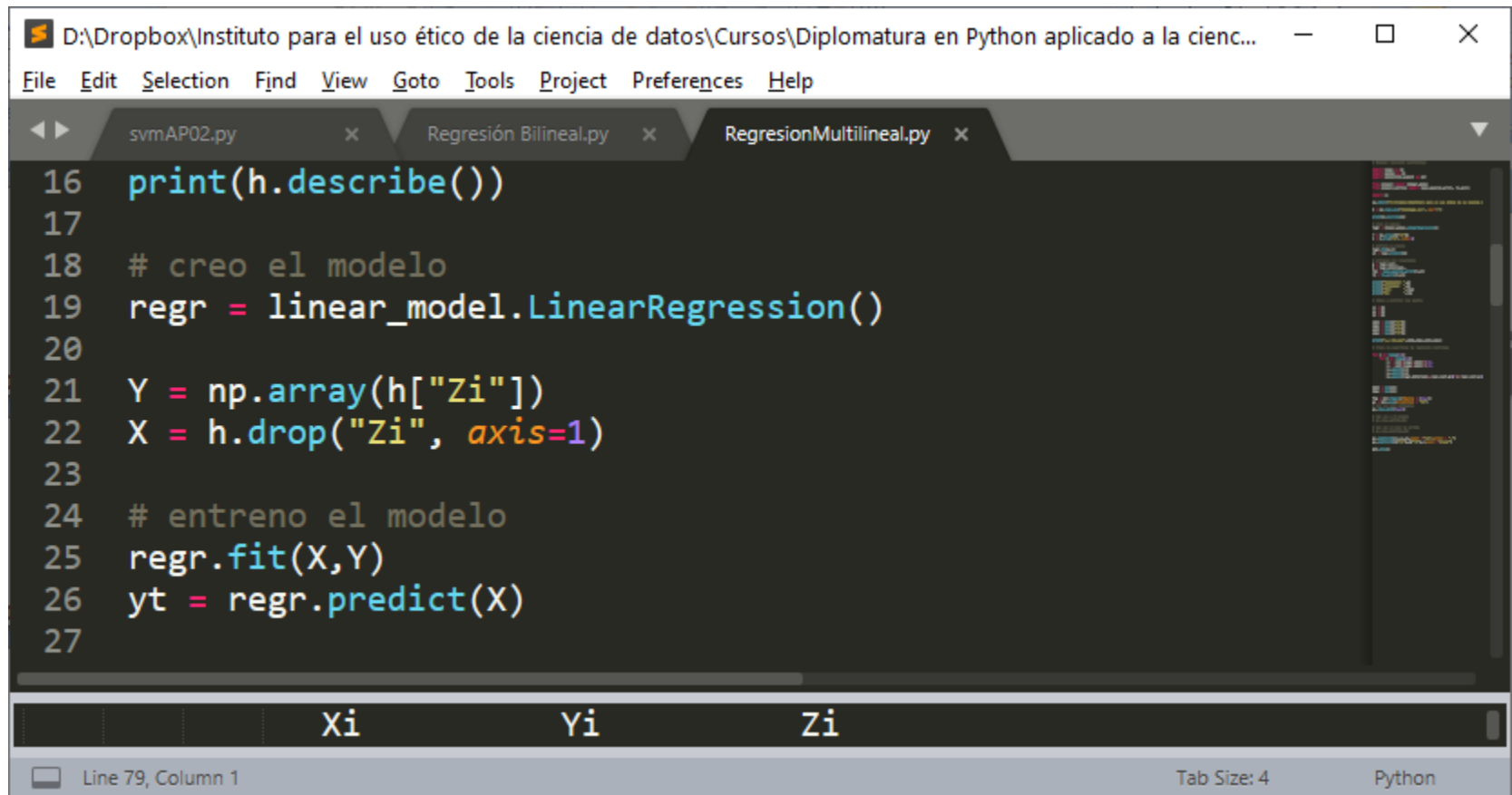
The image shows a screenshot of a Python IDE window. The title bar indicates the file path: `D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia...`. The menu bar includes `File`, `Edit`, `Selection`, `Find`, `View`, `Goto`, `Tools`, `Project`, `Preferences`, and `Help`. The tab bar shows three open files: `svmAP02.py`, `Regresión Bilineal.py`, and `RegresionMultilineal.py`. The active file, `RegresionMultilineal.py`, contains the following Python code:

```
1 # Ejemplo regresión multilineal
2
3 import numpy as np
4 import pandas as pd
5 import matplotlib.pyplot as plt
6
7 from sklearn import linear_model
8 from sklearn.metrics import mean_squared_error, r2_score
9
10 import os
11
12 os.chdir("D:/Dropbox/Instituto para el uso ético de la ciencia
13
14 h = pd.read_csv("bilineal.csv", sep=";")
```

Below the code editor, there is a table header with three columns: `Xi`, `Yi`, and `Zi`. The status bar at the bottom shows `Line 79, Column 1`, `Tab Size: 4`, and `Python`.



Ejemplo de regresión multilineal



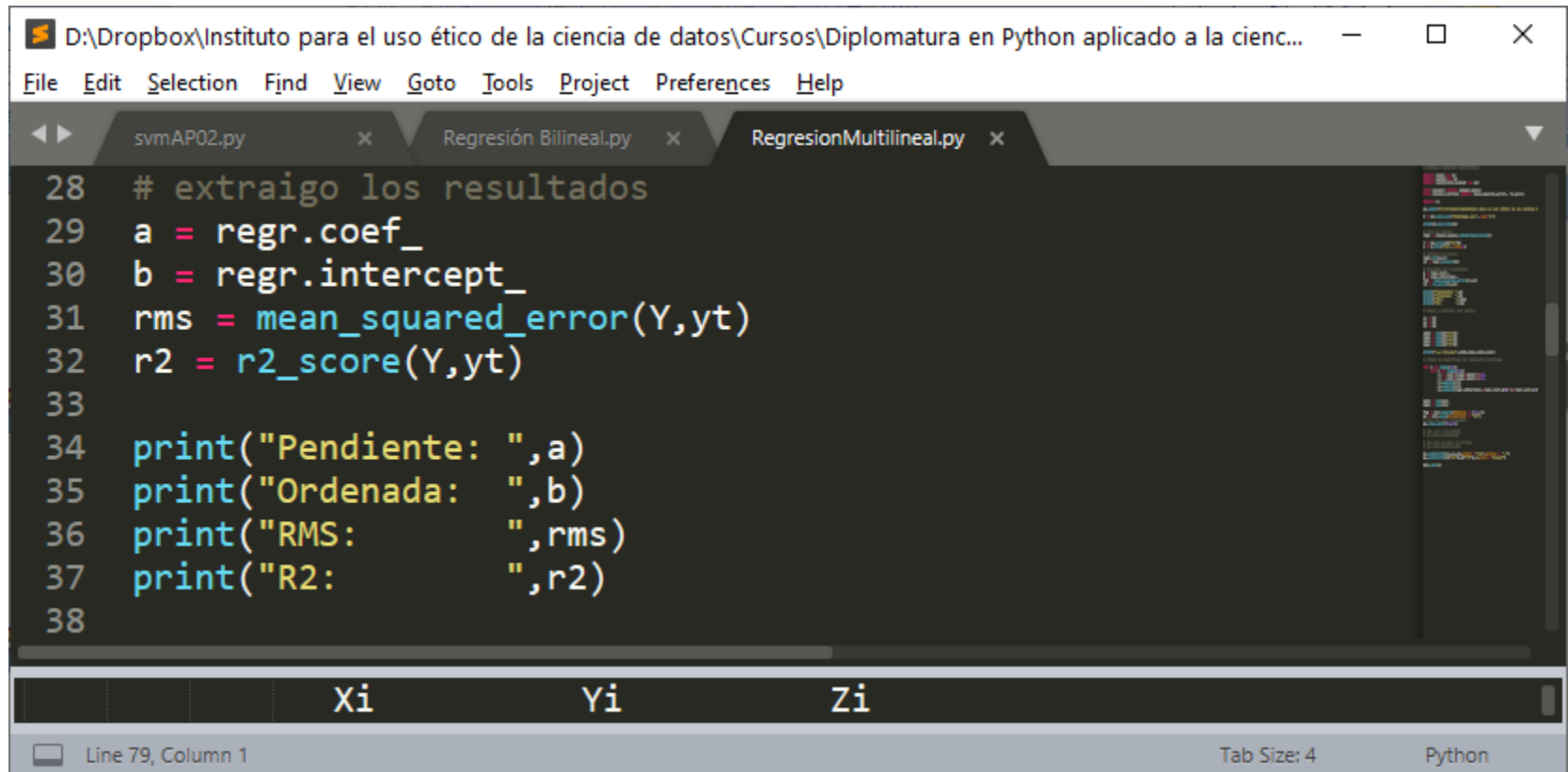
The screenshot shows a Python IDE window with the title bar "D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia...". The menu bar includes File, Edit, Selection, Find, View, Goto, Tools, Project, Preferences, and Help. The tab bar shows three open files: svmAP02.py, Regresión Bilineal.py, and RegresiónMultilineal.py. The main editor displays the following Python code:

```
16 print(h.describe())
17
18 # creo el modelo
19 regr = linear_model.LinearRegression()
20
21 Y = np.array(h["Zi"])
22 X = h.drop("Zi", axis=1)
23
24 # entreno el modelo
25 regr.fit(X,Y)
26 yt = regr.predict(X)
27
```

Below the code editor, there is a table with three columns labeled X_i , Y_i , and Z_i . The status bar at the bottom indicates "Line 79, Column 1", "Tab Size: 4", and "Python".



Ejemplo de regresión multilineal



The screenshot shows a Python IDE window titled "D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia...". The window contains three tabs: "svmAP02.py", "Regresión Bilineal.py", and "RegresionMultilineal.py". The active tab, "RegresionMultilineal.py", displays the following Python code:

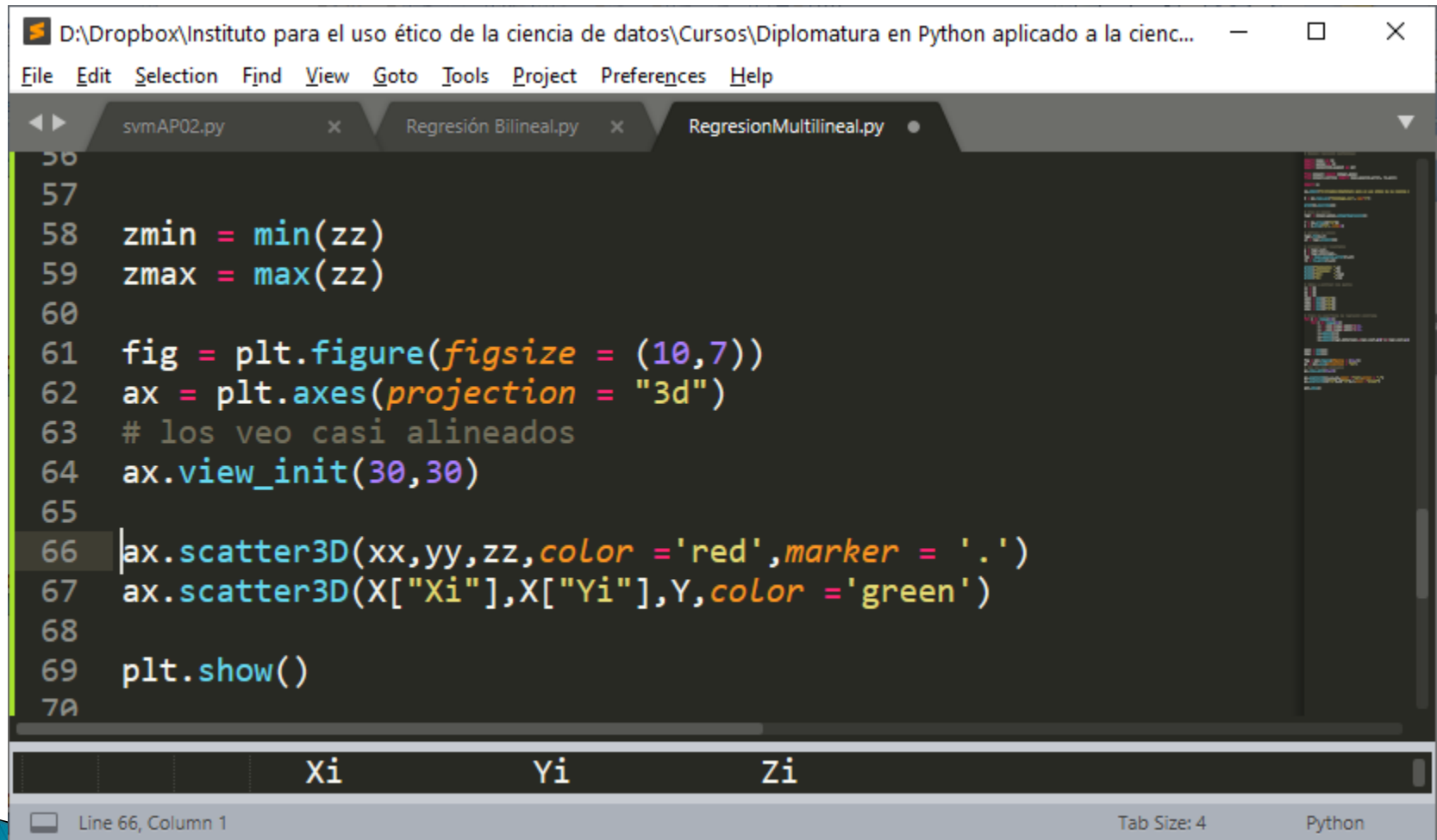
```
28 # extraigo los resultados
29 a = regr.coef_
30 b = regr.intercept_
31 rms = mean_squared_error(Y,yt)
32 r2 = r2_score(Y,yt)
33
34 print("Pendiente: ",a)
35 print("Ordenada: ",b)
36 print("RMS:      ",rms)
37 print("R2:       ",r2)
38
```

Below the code editor, there is a table with three columns labeled X_i , Y_i , and Z_i . The table is currently empty. The status bar at the bottom indicates "Line 79, Column 1", "Tab Size: 4", and "Python".



Ejemplo de regresión multilineal

Ejemplo de regresión multilineal



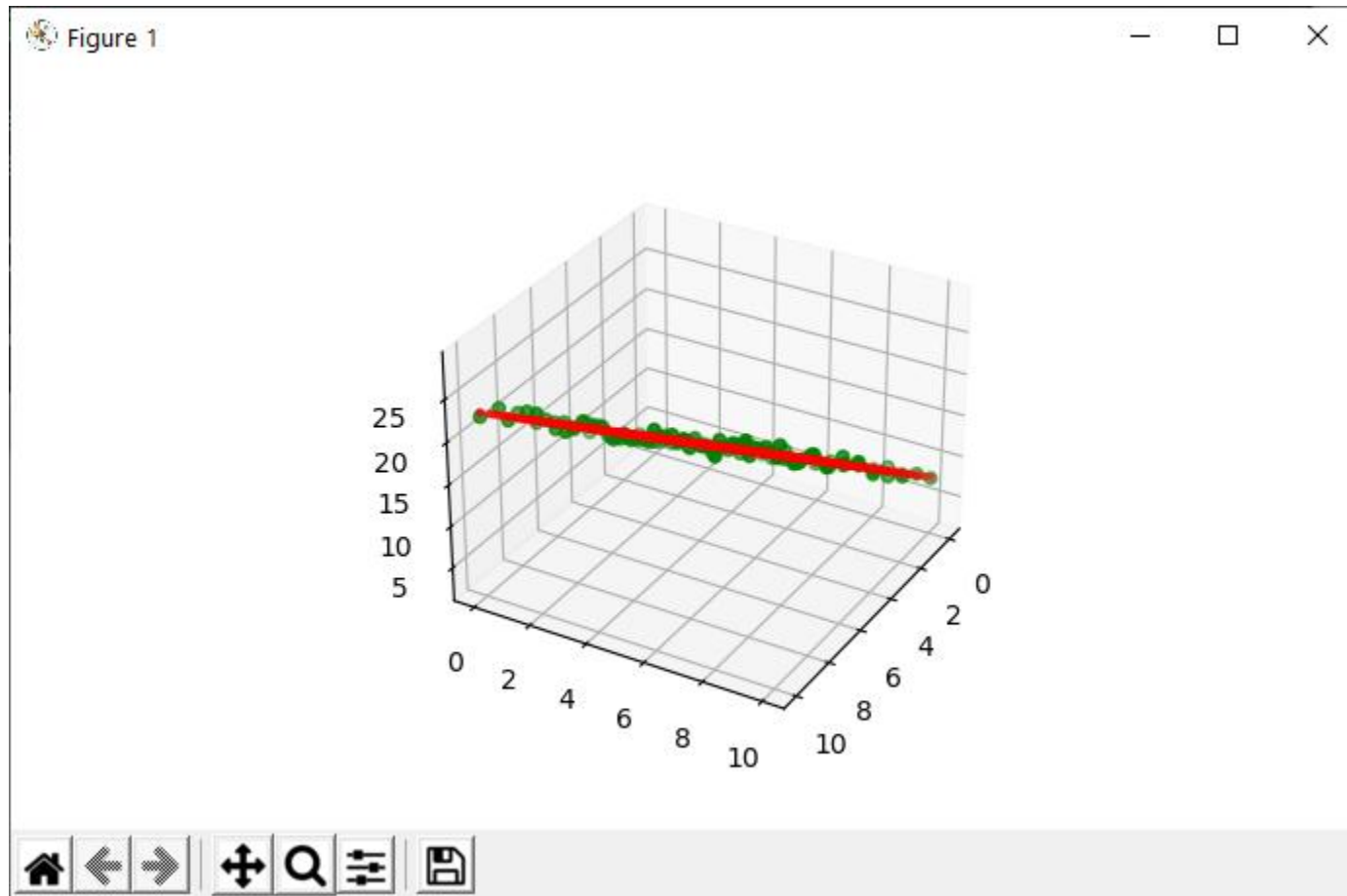
The image shows a screenshot of a Python IDE window. The title bar indicates the file path: `D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia...`. The menu bar includes `File`, `Edit`, `Selection`, `Find`, `View`, `Goto`, `Tools`, `Project`, `Preferences`, and `Help`. The tab bar shows three open files: `svmAP02.py`, `Regresión Bilineal.py`, and `RegresionMultilineal.py`. The main editor displays the following Python code:

```
56
57
58 zmin = min(zz)
59 zmax = max(zz)
60
61 fig = plt.figure(figsize = (10,7))
62 ax = plt.axes(projection = "3d")
63 # los veo casi alineados
64 ax.view_init(30,30)
65
66 ax.scatter3D(xx,yy,zz,color='red',marker='.')
67 ax.scatter3D(X["Xi"],X["Yi"],Y,color='green')
68
69 plt.show()
70
```

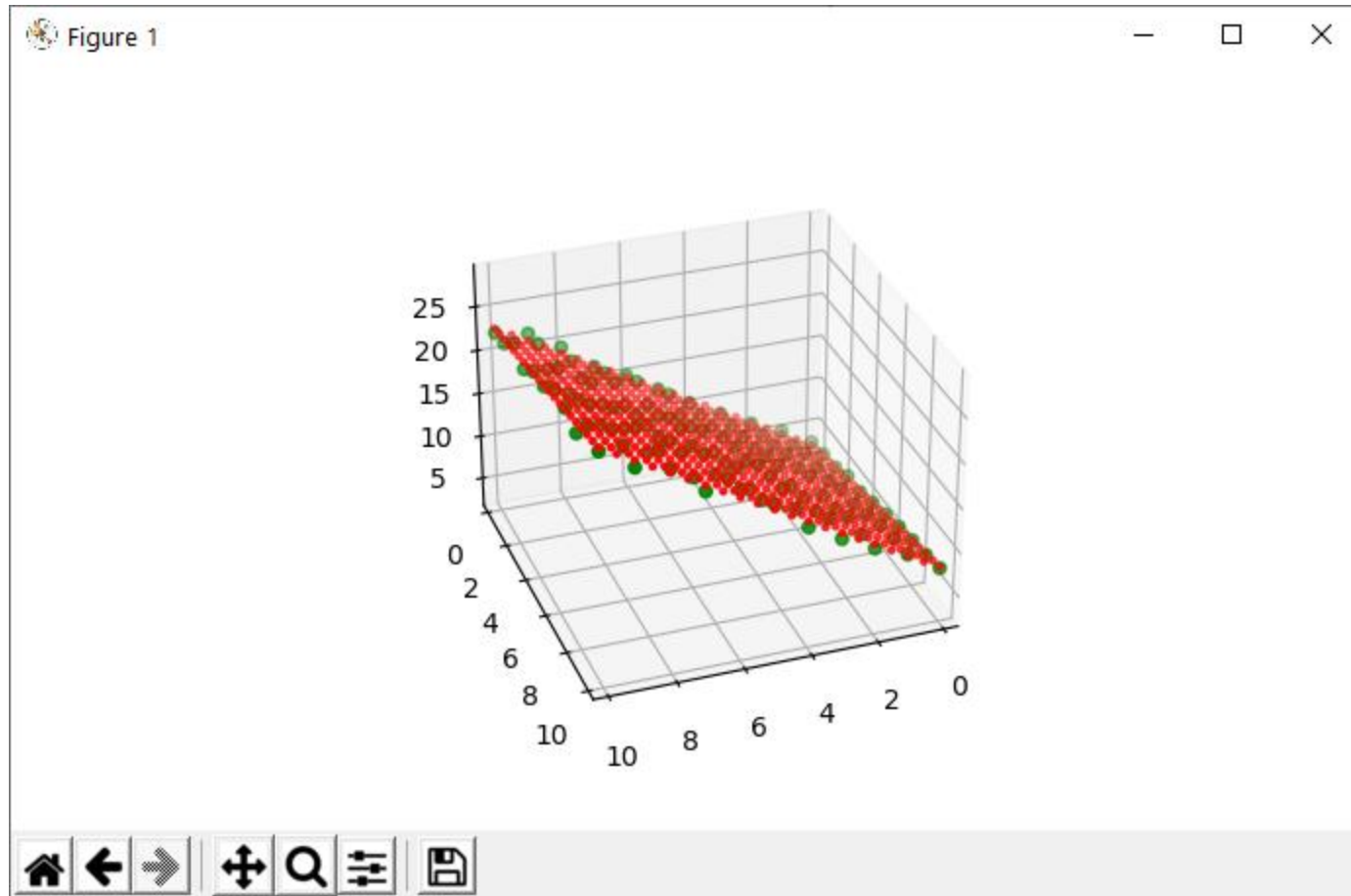
Below the code editor, there is a table with three columns labeled `Xi`, `Yi`, and `Zi`. The status bar at the bottom indicates `Line 66, Column 1`, `Tab Size: 4`, and `Python`.



Ejemplo de regresión multilineal



Ejemplo de regresión multilineal



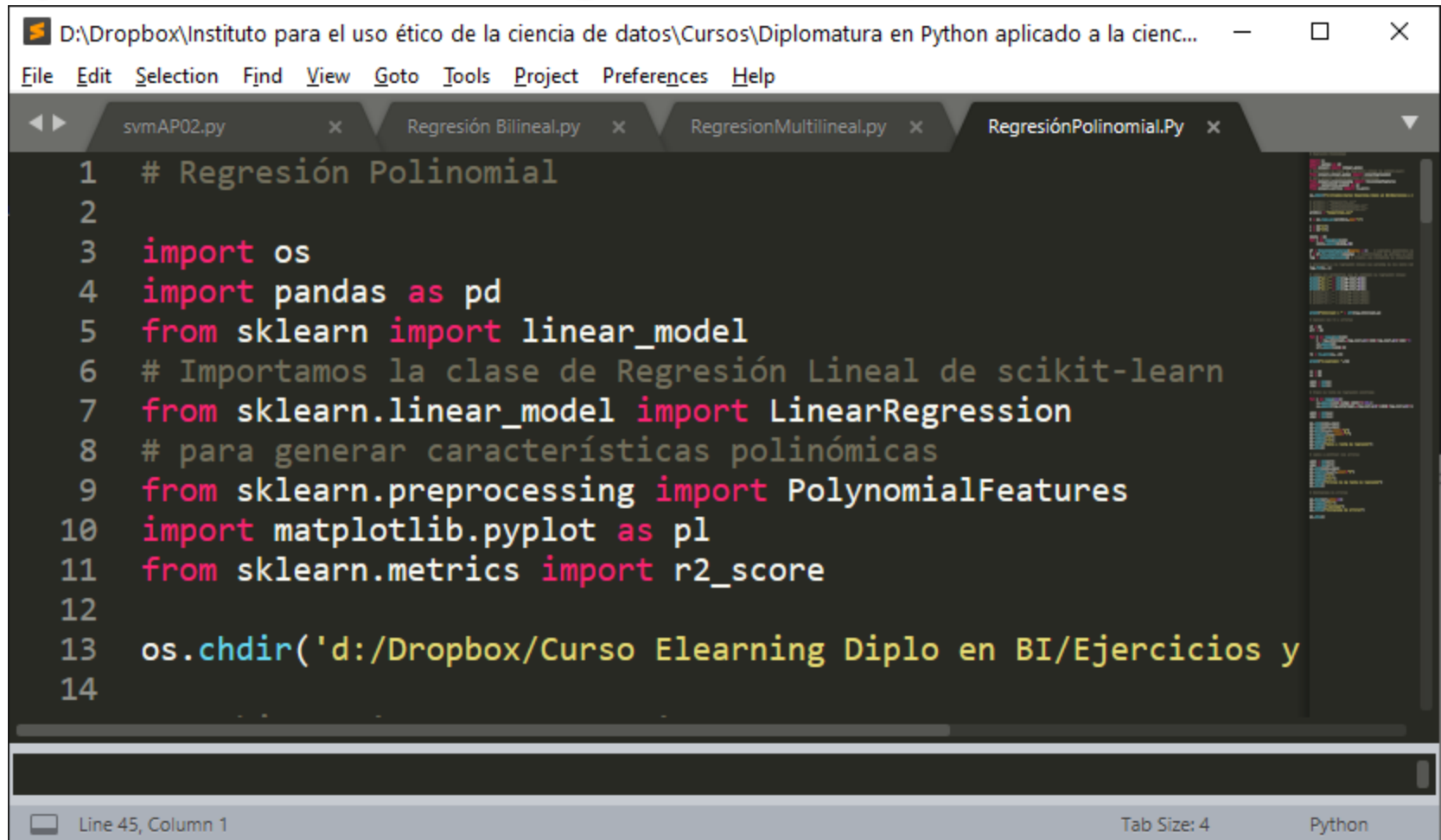
Ejemplo de regresión multilineal

```
D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia...
File Edit Selection Find View Goto Tools Project Preferences Help
svmAP02.py x Regresión Bilineal.py x RegresionMultilineal.py x
1 # Ejemplo regresión multilineal
2
count      Xi      Yi      Zi
mean       5.000000  5.000000  15.469450
std        3.175426  3.175426  6.626349
min        0.000000  0.000000  3.113543
25%        2.000000  2.000000  9.934891
50%        5.000000  5.000000  15.338393
75%        8.000000  8.000000  21.115212
max        10.000000 10.000000 28.591505
Pendiente: [2.01698599 0.51273145]
Ordenada:  2.8208630581983485
RMS:       0.23436699530830357
R2:        0.9946178982858961
x.v min.max 0 10 0 10
```

Line 79, Column 1 Tab Size: 4 Python



Ejemplo de regresión polinomial



The image shows a screenshot of a code editor window. The title bar indicates the file path: D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia... The editor has four tabs open: svmAP02.py, Regresión Bilineal.py, RegresiónMultilineal.py, and RegresiónPolinomial.py. The active tab is RegresiónPolinomial.py. The code is written in Python and includes comments in Spanish. The code imports necessary libraries (os, pandas, sklearn, matplotlib) and sets the working directory. The code is as follows:

```
1  # Regresión Polinomial
2
3  import os
4  import pandas as pd
5  from sklearn import linear_model
6  # Importamos la clase de Regresión Lineal de scikit-learn
7  from sklearn.linear_model import LinearRegression
8  # para generar características polinómicas
9  from sklearn.preprocessing import PolynomialFeatures
10 import matplotlib.pyplot as plt
11 from sklearn.metrics import r2_score
12
13 os.chdir('d:/Dropbox/Curso Elearning Diplo en BI/Ejercicios y
14
```

The status bar at the bottom shows "Line 45, Column 1", "Tab Size: 4", and "Python".



Ejemplo de regresión polinomial



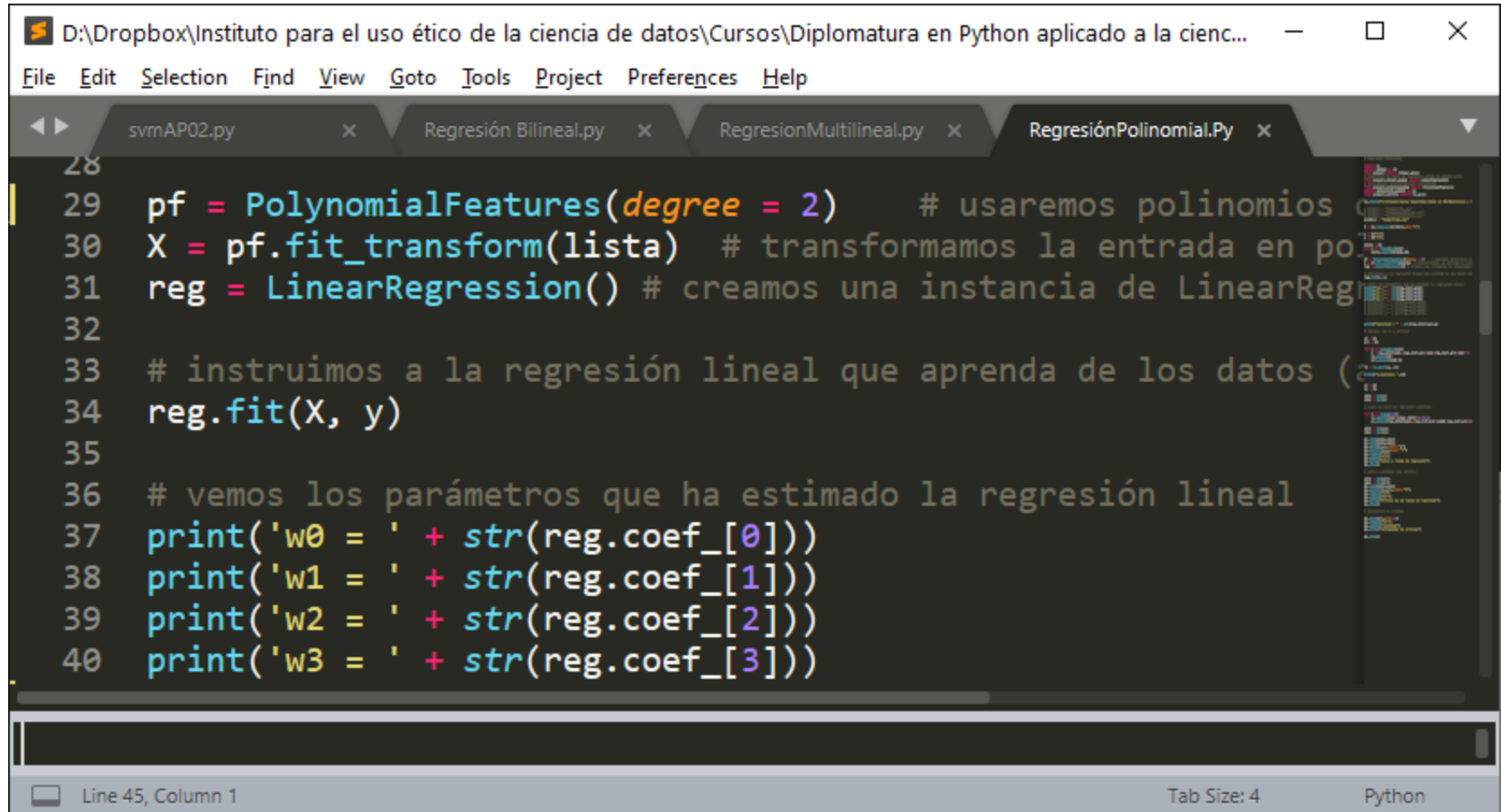
The image shows a screenshot of a Python IDE window. The title bar indicates the file path: `D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia de datos\...`. The menu bar includes `File`, `Edit`, `Selection`, `Find`, `View`, `Goto`, `Tools`, `Project`, `Preferences`, and `Help`. The tab bar shows four open files: `svmAP02.py`, `Regresión Bilineal.py`, `RegresiónMultilineal.py`, and `RegresiónPolinomial.py`. The active file, `RegresiónPolinomial.py`, contains the following Python code:

```
14
15 # archivo = 'RangoCorto.csv'
16 # archivo = 'RangoMedioInicial.csv'
17 # archivo = 'RangoMedioCentral.csv'
18 archivo = 'RangoTotal.csv'
19
20 h = pd.read_csv(archivo, sep=";")
21
22 x = h["Xi"]
23 y = h["Yi"]
24
25 lista = []
26 for i in range(len(x)):
27     lista.append([x[i], 0])
```

The status bar at the bottom shows `Line 45, Column 1`, `Tab Size: 4`, and `Python`.



Ejemplo de regresión polinomial




The screenshot shows a Python IDE window with the title bar "D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia...". The menu bar includes "File", "Edit", "Selection", "Find", "View", "Goto", "Tools", "Project", "Preferences", and "Help". The tab bar shows four open files: "svmAP02.py", "Regresión Bilineal.py", "RegresiónMultilineal.py", and "RegresiónPolinomial.py". The code editor displays the following Python code:

```
28
29 pf = PolynomialFeatures(degree = 2) # usaremos polinomios
30 X = pf.fit_transform(lista) # transformamos la entrada en po
31 reg = LinearRegression() # creamos una instancia de LinearReg
32
33 # instruimos a la regresión lineal que aprenda de los datos (
34 reg.fit(X, y)
35
36 # vemos los parámetros que ha estimado la regresión lineal
37 print('w0 = ' + str(reg.coef_[0]))
38 print('w1 = ' + str(reg.coef_[1]))
39 print('w2 = ' + str(reg.coef_[2]))
40 print('w3 = ' + str(reg.coef_[3]))
```

The status bar at the bottom indicates "Line 45, Column 1", "Tab Size: 4", and "Python".



Ejemplo de regresión polinomial



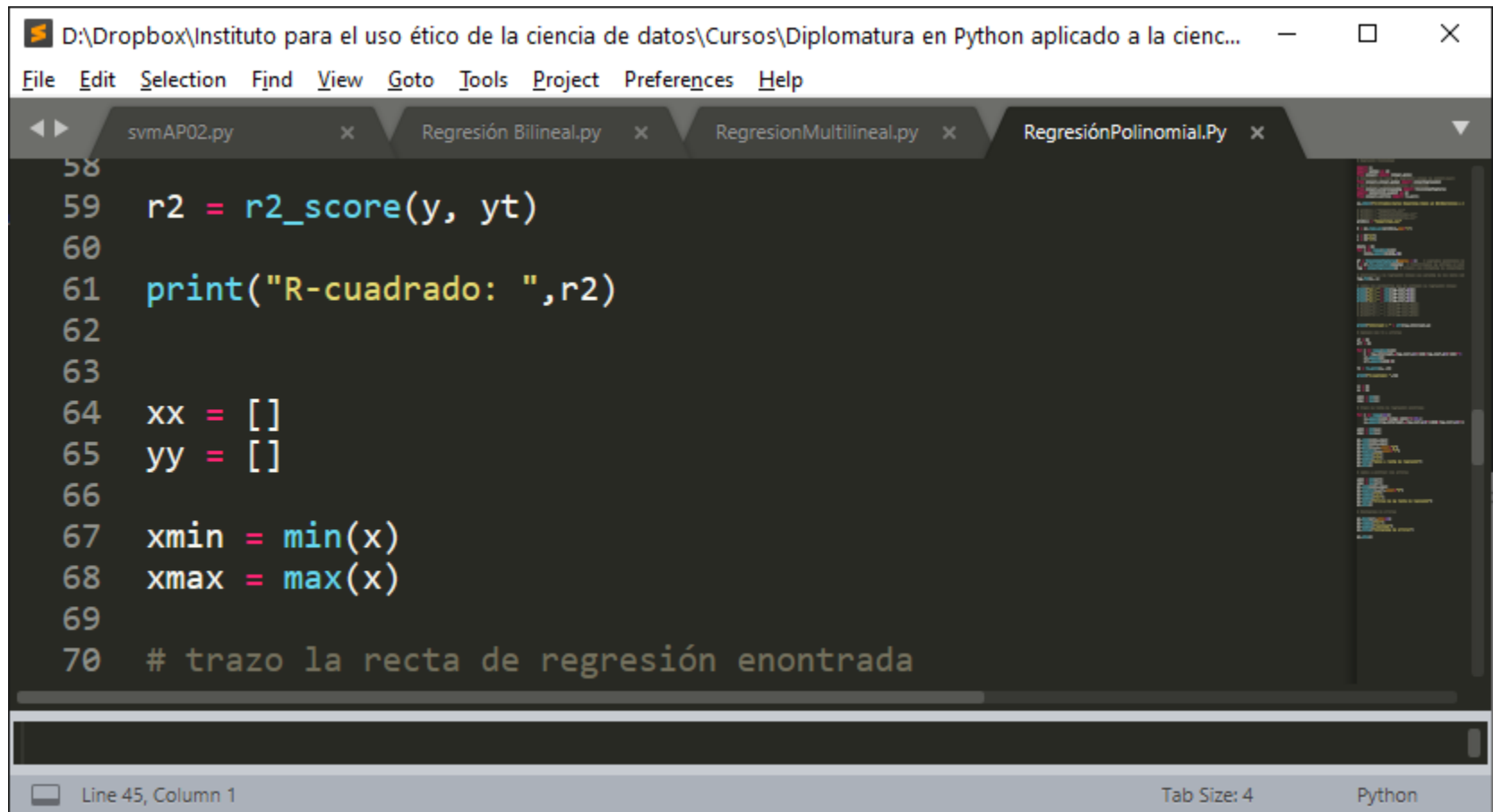
The screenshot shows a Python IDE window with the title bar "D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la cienc...". The menu bar includes File, Edit, Selection, Find, View, Goto, Tools, Project, Preferences, and Help. The tab bar shows four open files: svmAP02.py, Regresión Bilineal.py, RegresiónMultilineal.py, and RegresiónPolinomial.py. The active file, RegresiónPolinomial.py, contains the following Python code:

```
46
47 print('intercept = ' + str(reg.intercept_))
48
49 # Calculo del r2 y errores
50
51 yt = []
52 err = []
53
54 for i in range(len(x)):
55     b = reg.intercept_+reg.coef_[1]*x[i]+reg.coef_[3]*x[i]**2
56     yt.append(b)
57     err.append(y[i]-b)
58
```

The status bar at the bottom indicates "Line 45, Column 1", "Tab Size: 4", and "Python".



Ejemplo de regresión polinomial



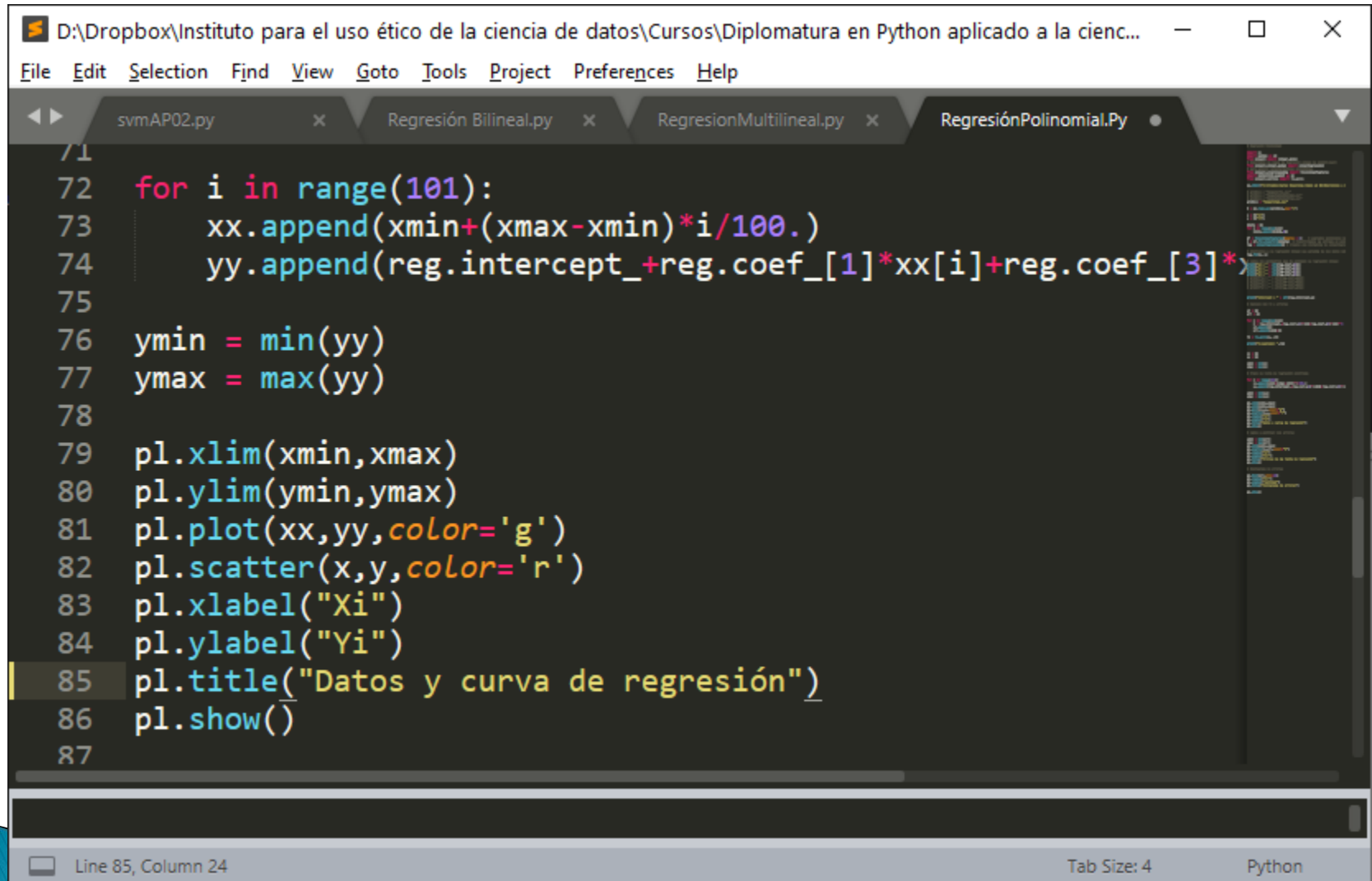
The image shows a screenshot of a Python IDE window. The title bar indicates the file path: D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia... The menu bar includes File, Edit, Selection, Find, View, Goto, Tools, Project, Preferences, and Help. The tab bar shows four open files: svmAP02.py, Regresión Bilineal.py, RegresiónMultilineal.py, and RegresiónPolinomial.Py. The main editor area displays the following Python code:

```
58
59 r2 = r2_score(y, yt)
60
61 print("R-cuadrado: ",r2)
62
63
64 xx = []
65 yy = []
66
67 xmin = min(x)
68 xmax = max(x)
69
70 # trazo la recta de regresión enontrada
```

The status bar at the bottom shows "Line 45, Column 1", "Tab Size: 4", and "Python".



Ejemplo de regresión polinomial



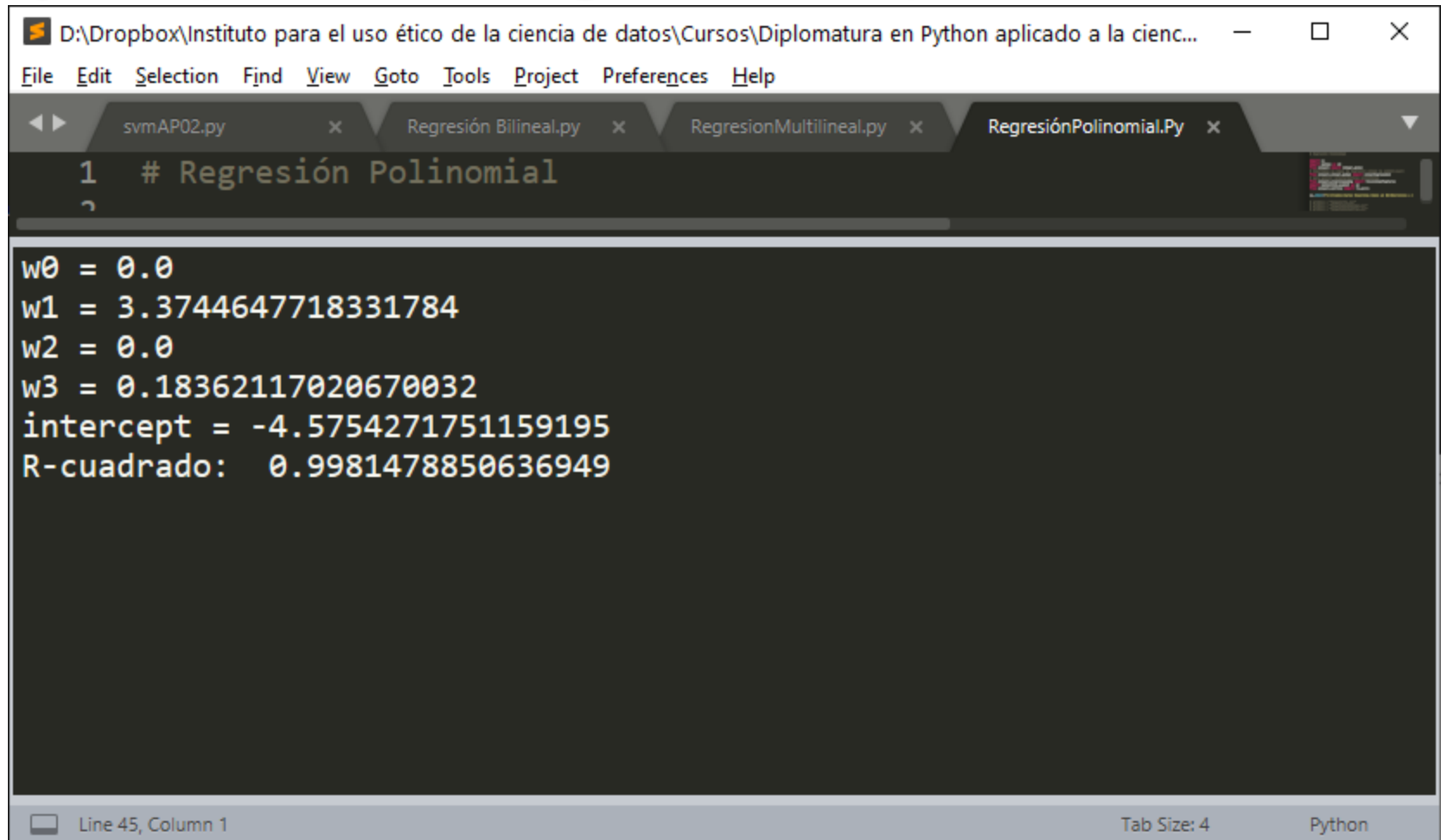
The image shows a screenshot of a Python IDE window. The title bar indicates the file path: D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia... The menu bar includes File, Edit, Selection, Find, View, Goto, Tools, Project, Preferences, and Help. The tab bar shows four open files: svmAP02.py, Regresión Bilineal.py, RegresiónMultilineal.py, and RegresiónPolinomial.py. The main editor displays the following Python code:

```
71
72 for i in range(101):
73     xx.append(xmin+(xmax-xmin)*i/100.)
74     yy.append(reg.intercept_+reg.coef_[1]*xx[i]+reg.coef_[3]*>
75
76 ymin = min(yy)
77 ymax = max(yy)
78
79 pl.xlim(xmin,xmax)
80 pl.ylim(ymin,ymax)
81 pl.plot(xx,yy,color='g')
82 pl.scatter(x,y,color='r')
83 pl.xlabel("Xi")
84 pl.ylabel("Yi")
85 pl.title("Datos y curva de regresión")
86 pl.show()
87
```

The status bar at the bottom shows "Line 85, Column 24", "Tab Size: 4", and "Python".



Ejemplo de regresión polinomial



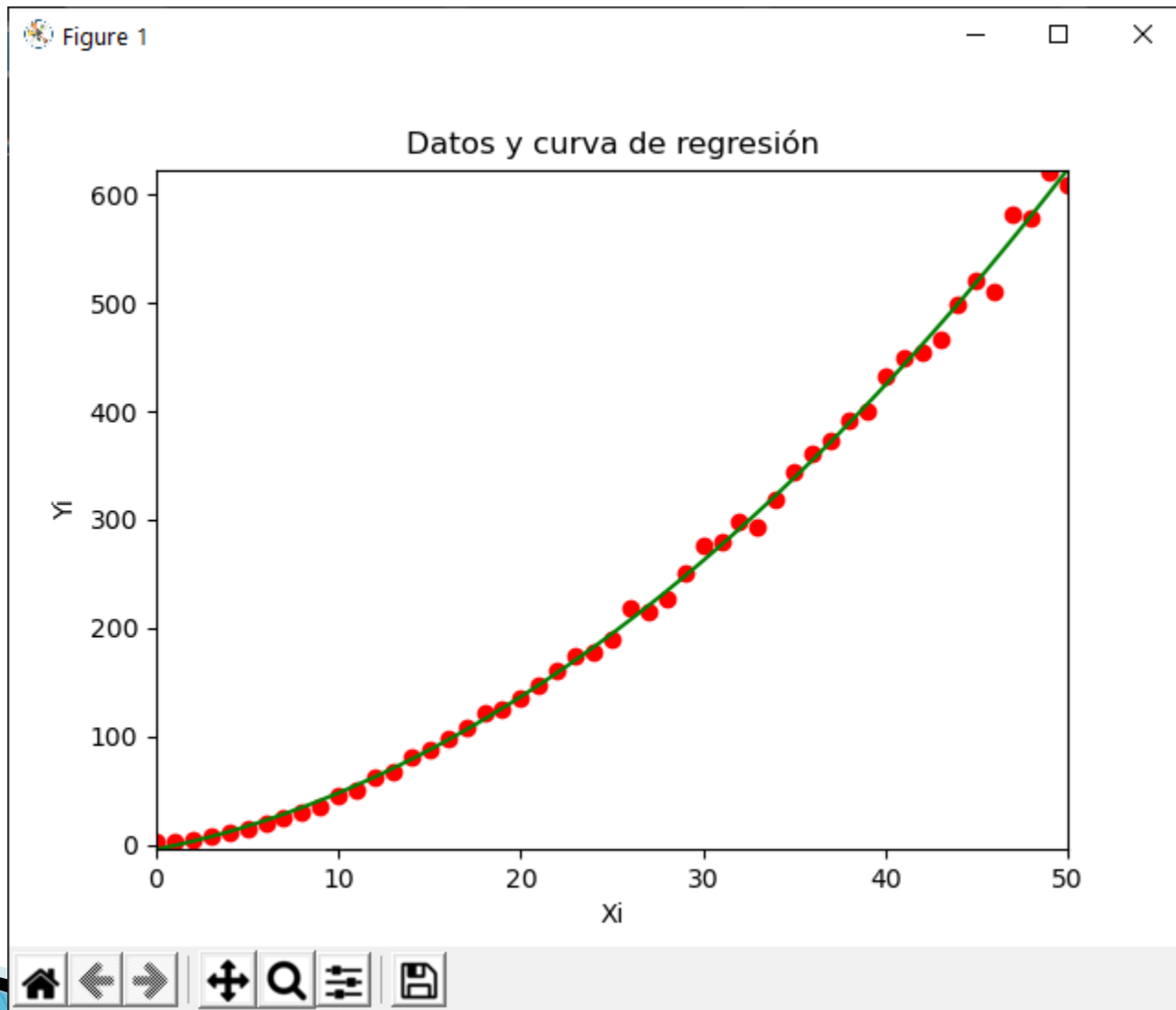
The image shows a screenshot of a Python IDE window. The title bar indicates the file path: D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia... The menu bar includes File, Edit, Selection, Find, View, Goto, Tools, Project, Preferences, and Help. The tab bar shows four open files: svmAP02.py, Regresión Bilineal.py, RegresiónMultilineal.py, and RegresiónPolinomial.py. The active file, RegresiónPolinomial.py, contains the following code:

```
1 # Regresión Polinomial
2
w0 = 0.0
w1 = 3.3744647718331784
w2 = 0.0
w3 = 0.18362117020670032
intercept = -4.5754271751159195
R-cuadrado: 0.9981478850636949
```

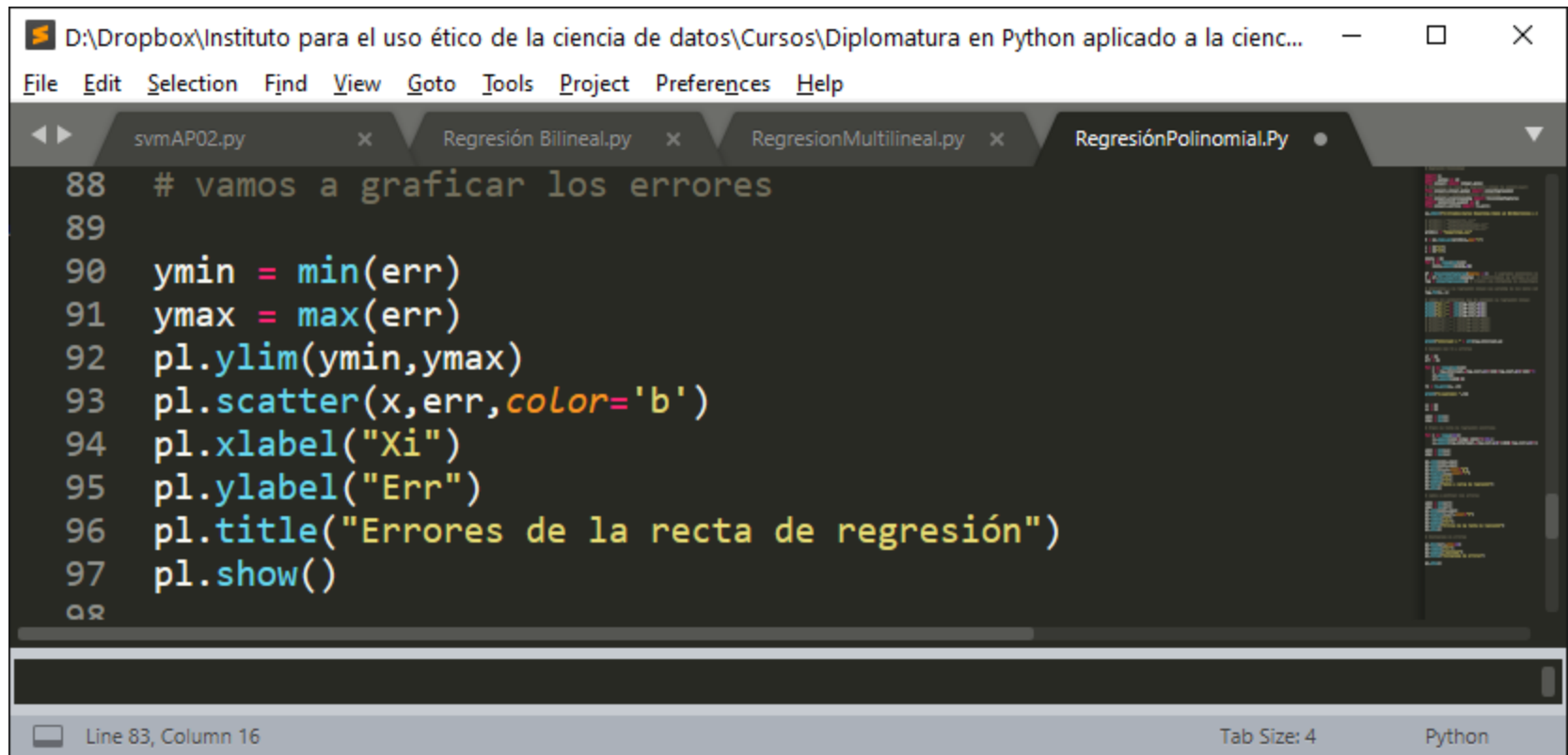
The status bar at the bottom shows "Line 45, Column 1", "Tab Size: 4", and "Python".



Ejemplo de regresión polinomial



Ejemplo de regresión polinomial



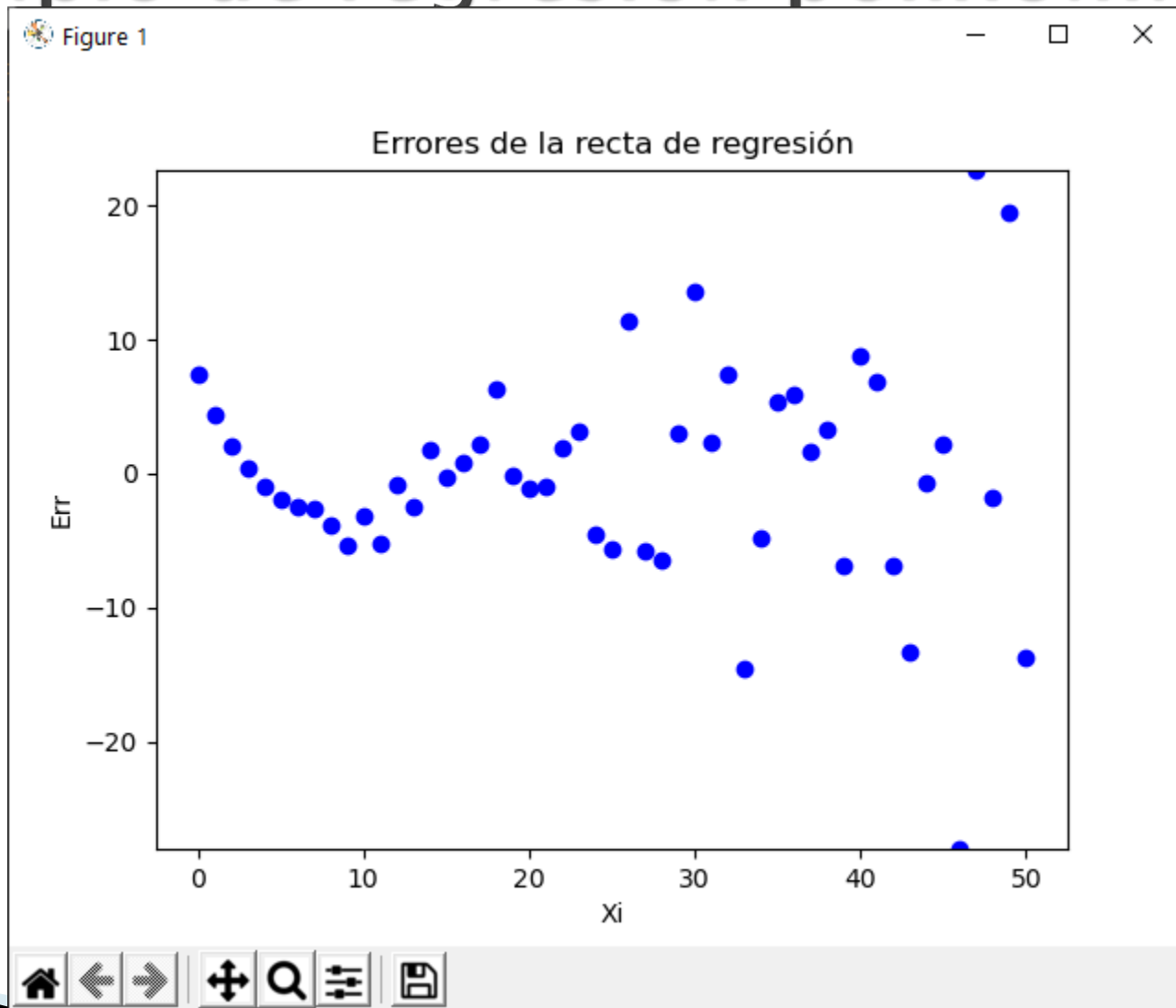
The image shows a screenshot of a Python IDE window. The title bar indicates the file path: `D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia de datos\`. The menu bar includes `File`, `Edit`, `Selection`, `Find`, `View`, `Goto`, `Tools`, `Project`, `Preferences`, and `Help`. The tab bar shows four open files: `svmAP02.py`, `Regresión Bilineal.py`, `RegresiónMultilineal.py`, and `RegresiónPolinomial.Py`. The active file, `RegresiónPolinomial.Py`, contains the following Python code:

```
88 # vamos a graficar los errores
89
90 ymin = min(err)
91 ymax = max(err)
92 pl.ylim(ymin,ymax)
93 pl.scatter(x,err,color='b')
94 pl.xlabel("Xi")
95 pl.ylabel("Err")
96 pl.title("Errores de la recta de regresión")
97 pl.show()
98
```

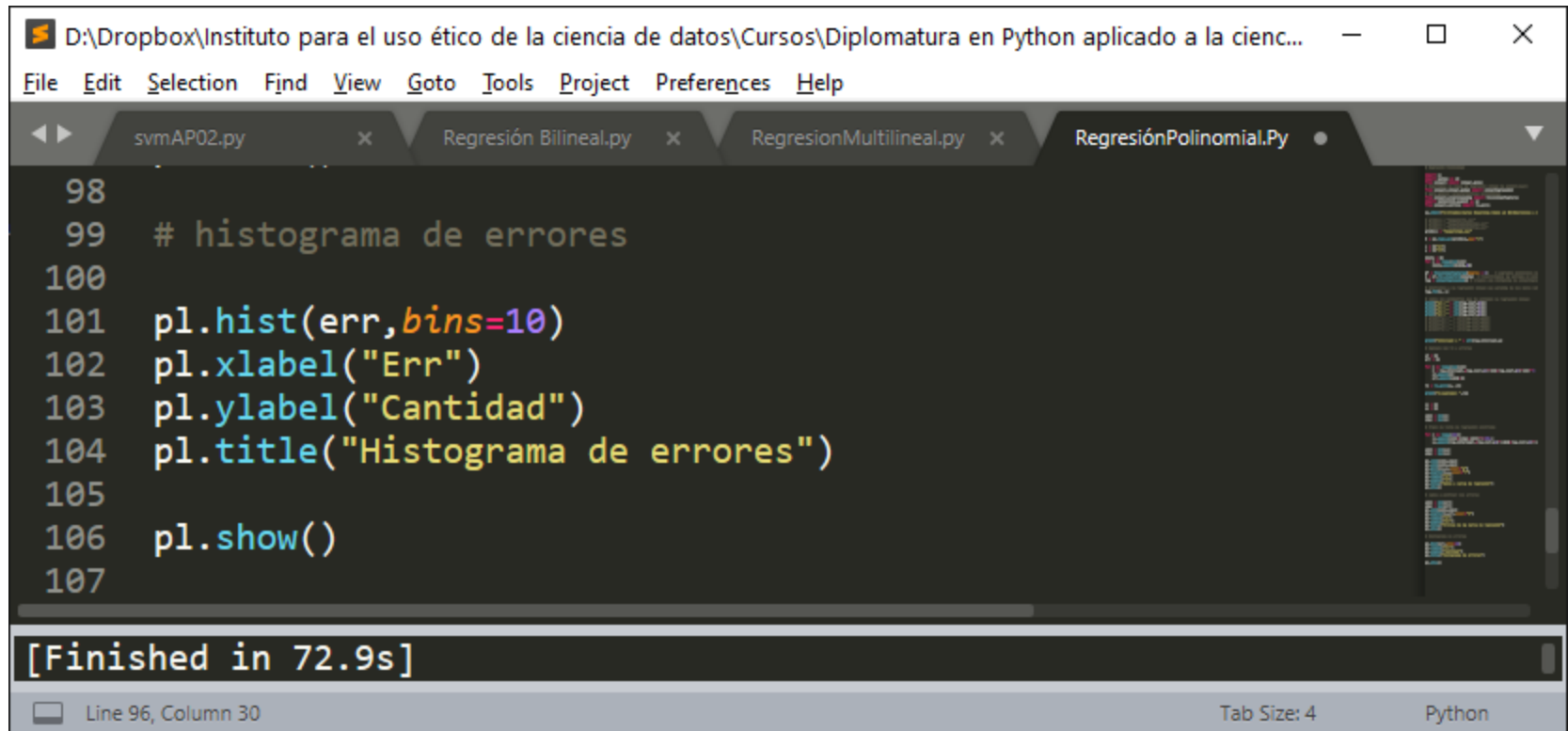
The status bar at the bottom shows the cursor position as `Line 83, Column 16`, the tab size as `Tab Size: 4`, and the language as `Python`.



Ejemplo de regresión polinomial



Ejemplo de regresión polinomial



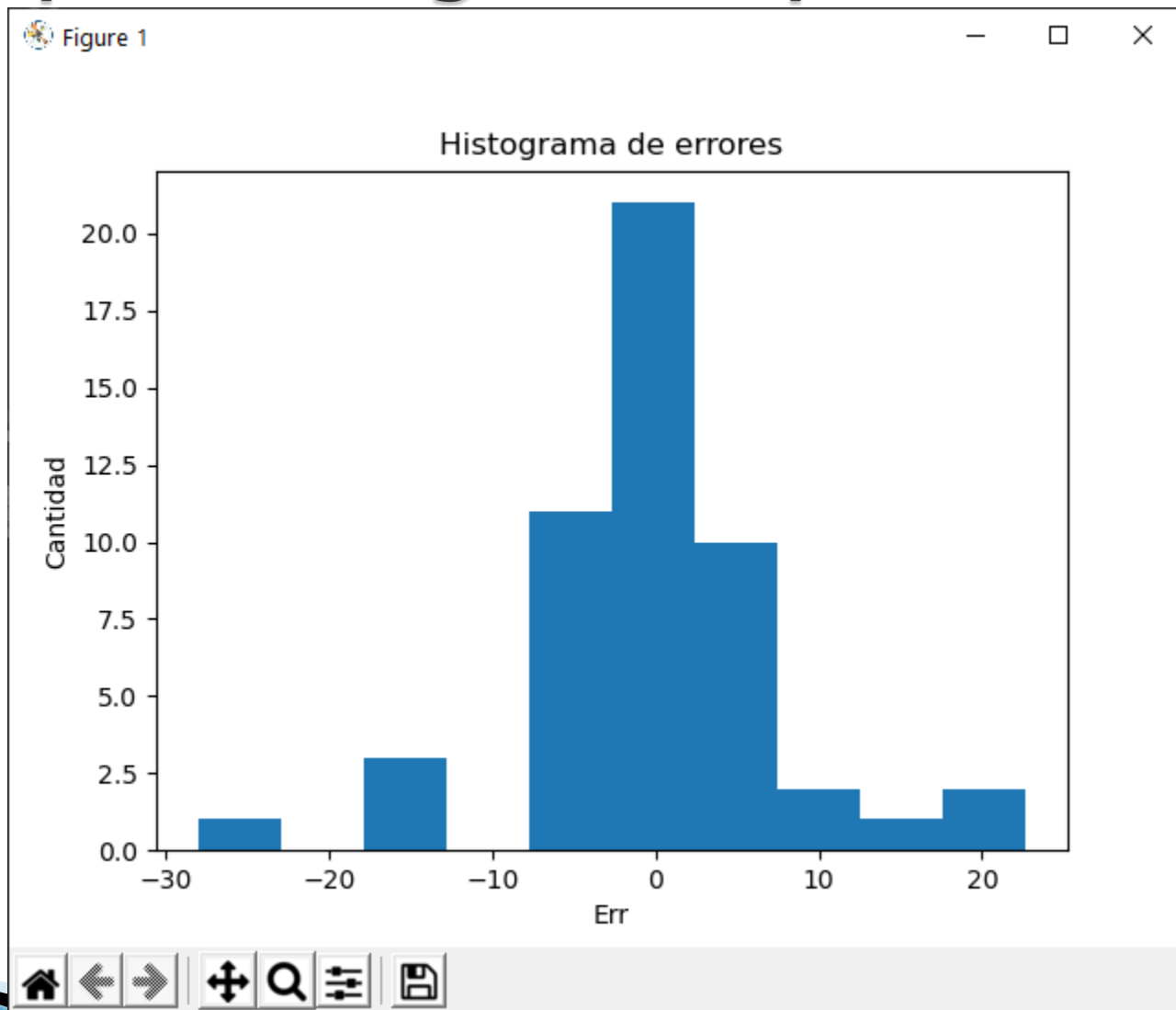
The screenshot shows a Python IDE window with the title bar "D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia...". The menu bar includes "File", "Edit", "Selection", "Find", "View", "Goto", "Tools", "Project", "Preferences", and "Help". The tab bar shows four open files: "svmAP02.py", "Regresión Bilineal.py", "RegresiónMultilineal.py", and "RegresiónPolinomial.py". The main editor displays the following Python code:

```
98
99 # histograma de errores
100
101 pl.hist(err,bins=10)
102 pl.xlabel("Err")
103 pl.ylabel("Cantidad")
104 pl.title("Histograma de errores")
105
106 pl.show()
107
```

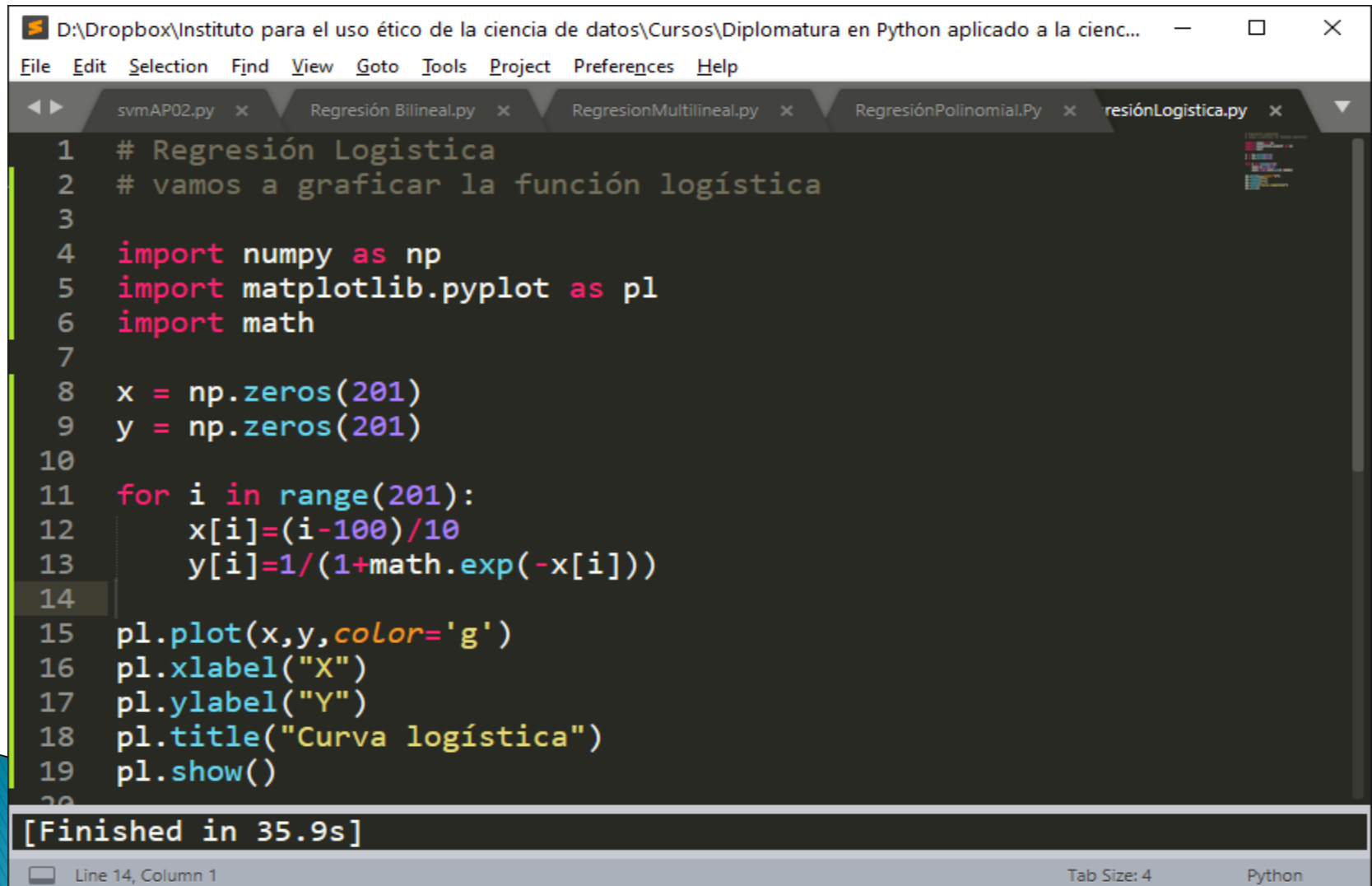
Below the code editor, a console window shows the output "[Finished in 72.9s]". The status bar at the bottom indicates "Line 96, Column 30", "Tab Size: 4", and "Python".



Ejemplo de regresión polinomial



Función logística

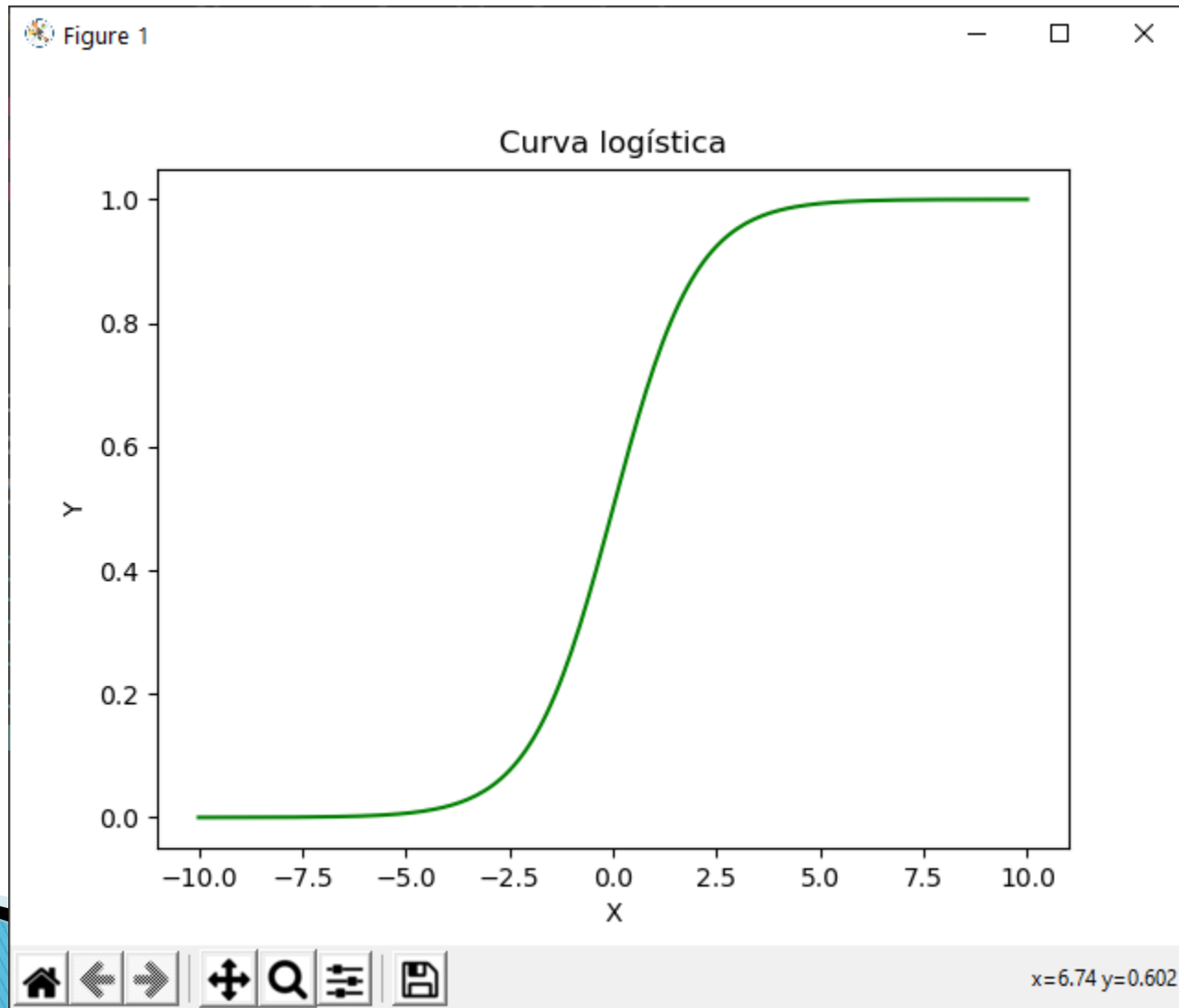


The image shows a screenshot of a Python IDE window. The title bar indicates the file path: D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia... The menu bar includes File, Edit, Selection, Find, View, Goto, Tools, Project, Preferences, and Help. The tab bar shows several open files: svmAP02.py, Regresión Bilineal.py, RegresiónMultilineal.py, RegresiónPolinomial.py, and the active file, resiónLogistica.py. The code editor displays the following Python code:

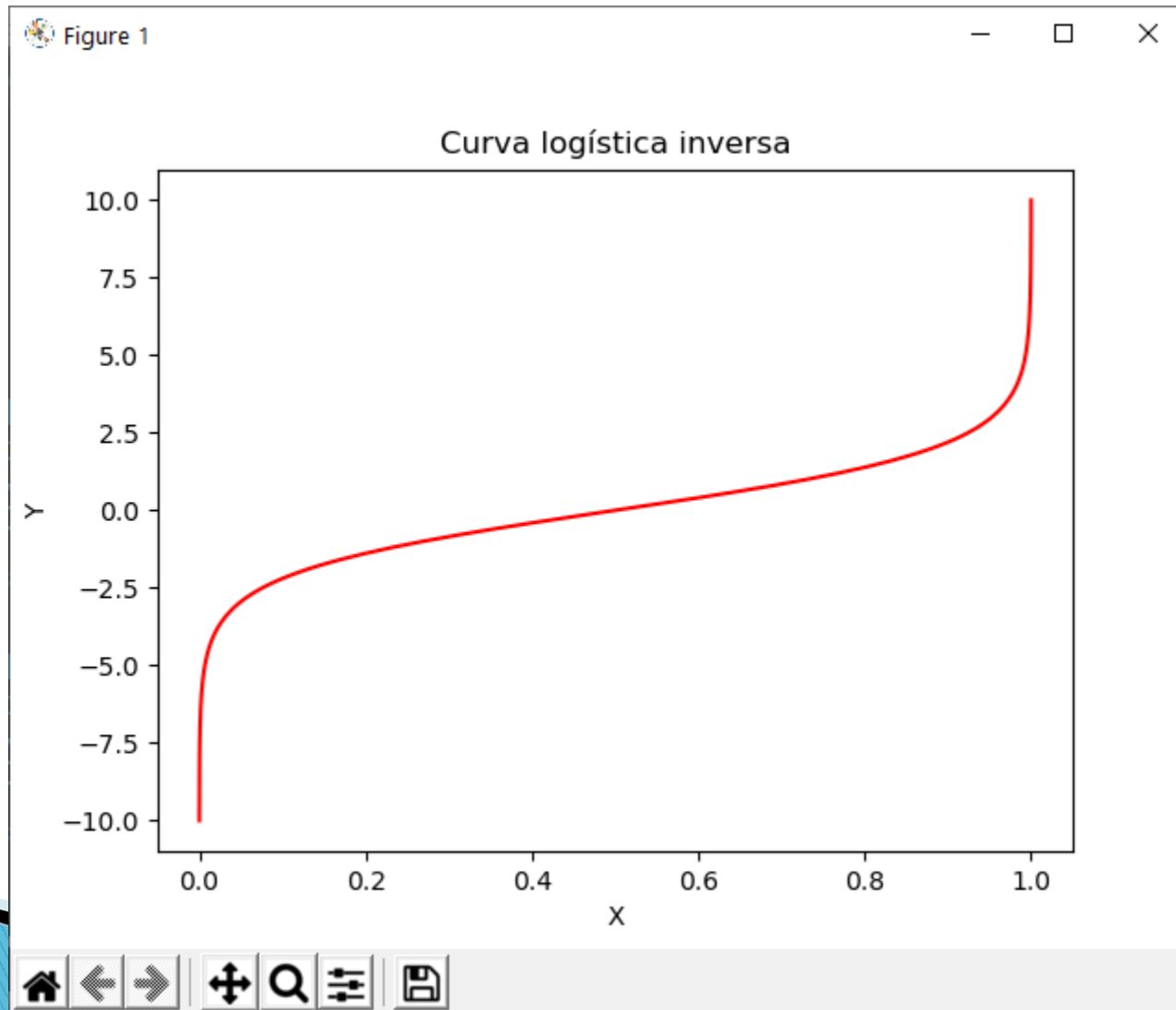
```
1 # Regresión Logistica
2 # vamos a graficar la función logística
3
4 import numpy as np
5 import matplotlib.pyplot as plt
6 import math
7
8 x = np.zeros(201)
9 y = np.zeros(201)
10
11 for i in range(201):
12     x[i]=(i-100)/10
13     y[i]=1/(1+math.exp(-x[i]))
14
15 plt.plot(x,y,color='g')
16 plt.xlabel("X")
17 plt.ylabel("Y")
18 plt.title("Curva logística")
19 plt.show()
20
```

At the bottom of the window, a status bar shows "[Finished in 35.9s]", "Line 14, Column 1", "Tab Size: 4", and "Python". A small globe icon is visible in the bottom right corner.

Función logística



Función logística inversa = logit



Regresión logística

Planteamos:

$$\text{Logit}(P_j) = C + \sum_{i=1}^N \beta_i X_i$$

Donde:

- P_i es la probabilidad de que un elemento pertenezca a la clase j
- X_i son los atributos numéricos del punto a clasificar
- Los β_i y C son los factores que vamos a buscar mediante regresión.



Regresión logística

Tomo $\text{logist}()$ en ambos miembros:

$$\text{Logist}(\text{Logit}(P_j)) = \text{logist}(C + \sum_{i=1}^N \beta_i X_i)$$

Pero:

- Logist y Logit son las inversas, al componerlas tengo la identidad:



Regresión logística

$$P_j = \text{logist} \left(C + \sum_{i=1}^N \beta_i X_i \right)$$

Reemplazo logist por su definición y a C lo llamo Beta0:

$$P_j = \frac{1}{1 + \exp \left(- \left(\beta_0 + \sum_{i=1}^N \beta_i X_i \right) \right)}$$



Regresión logística

Conocidos los Betas podremos calcular las probabilidades de que cada caso pertenezca a cada clase j .

Luego podremos asignar cada caso a la clase a la que su probabilidad sea máxima.



Regresión logística

¿Qué significan los Betas?

Beta_i mide cuanto “tira” X_i

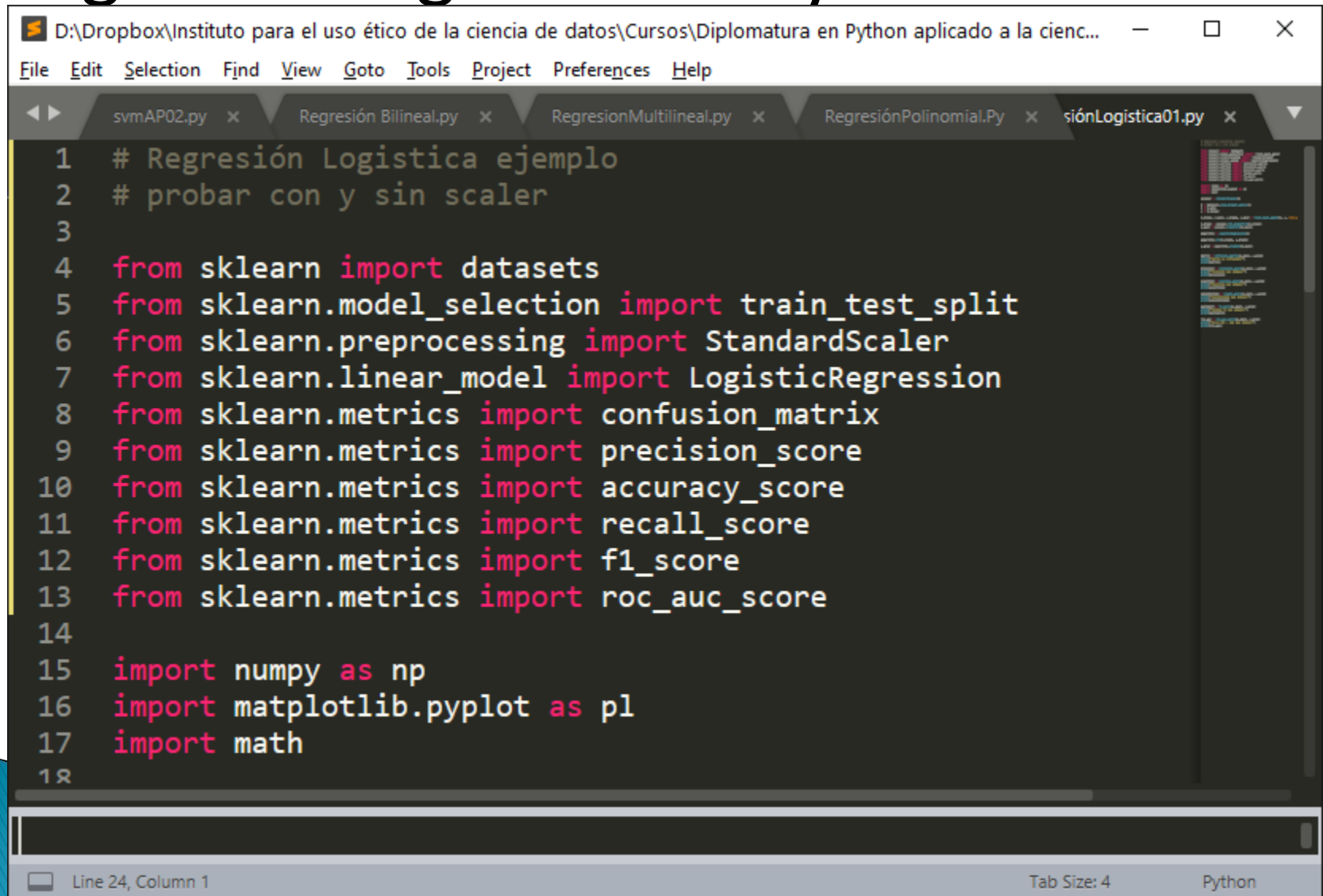
Si sólo tengo dos clases exp(Beta_i) es corresponde con el cociente entre las probabilidades de las dos clases que se conoce como Odds Ratio

Si, por ejemplo las dos clases son rubio o castaño entonces

$$\frac{P(\text{rubio})}{P(\text{castaño})} = \exp(\beta_i)$$



Regresión logística en Python

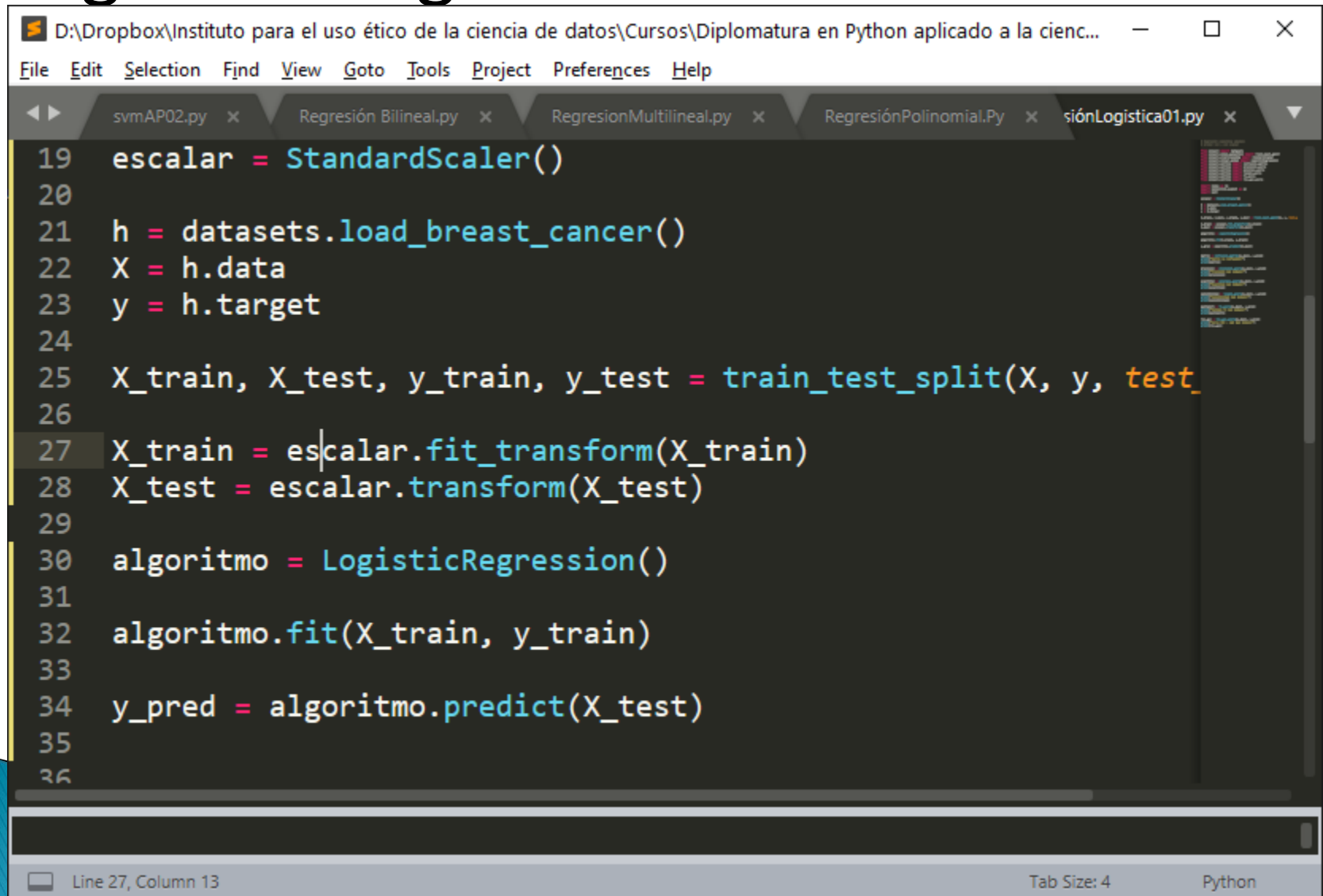


The image shows a screenshot of a Python IDE window. The title bar indicates the file path: D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia... The menu bar includes File, Edit, Selection, Find, View, Goto, Tools, Project, Preferences, and Help. The tab bar shows several open files: svmAP02.py, Regresión Bilineal.py, RegresiónMultilineal.py, RegresiónPolinomial.py, and siónLogistica01.py. The main editor area displays the following Python code:

```
1 # Regresión Logistica ejemplo
2 # probar con y sin scaler
3
4 from sklearn import datasets
5 from sklearn.model_selection import train_test_split
6 from sklearn.preprocessing import StandardScaler
7 from sklearn.linear_model import LogisticRegression
8 from sklearn.metrics import confusion_matrix
9 from sklearn.metrics import precision_score
10 from sklearn.metrics import accuracy_score
11 from sklearn.metrics import recall_score
12 from sklearn.metrics import f1_score
13 from sklearn.metrics import roc_auc_score
14
15 import numpy as np
16 import matplotlib.pyplot as plt
17 import math
18
```

The status bar at the bottom shows "Line 24, Column 1", "Tab Size: 4", and "Python". A small globe icon is visible in the bottom right corner.

Regresión logística

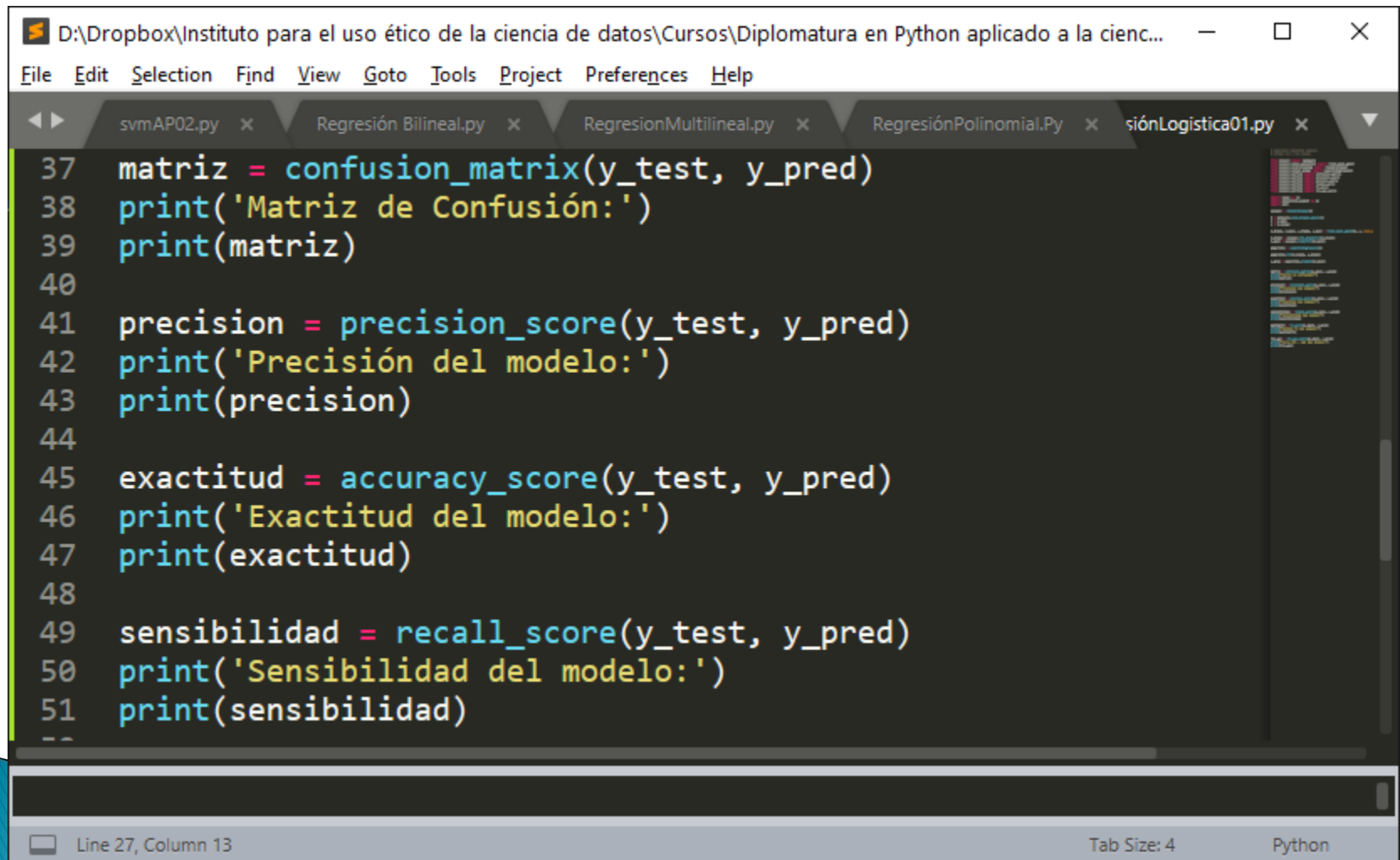


The screenshot shows a Python IDE window with the title bar "D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia...". The menu bar includes File, Edit, Selection, Find, View, Goto, Tools, Project, Preferences, and Help. The tab bar shows several open files: svmAP02.py, Regresión Bilineal.py, RegresiónMultilineal.py, RegresiónPolinomial.py, and siónLogistica01.py. The main editor area displays the following Python code:

```
19 escalar = StandardScaler()
20
21 h = datasets.load_breast_cancer()
22 X = h.data
23 y = h.target
24
25 X_train, X_test, y_train, y_test = train_test_split(X, y, test_
26
27 X_train = escalar.fit_transform(X_train)
28 X_test = escalar.transform(X_test)
29
30 algoritmo = LogisticRegression()
31
32 algoritmo.fit(X_train, y_train)
33
34 y_pred = algoritmo.predict(X_test)
35
36
```

The status bar at the bottom indicates "Line 27, Column 13", "Tab Size: 4", and "Python". A small globe icon is visible in the bottom right corner.

Regresión logística

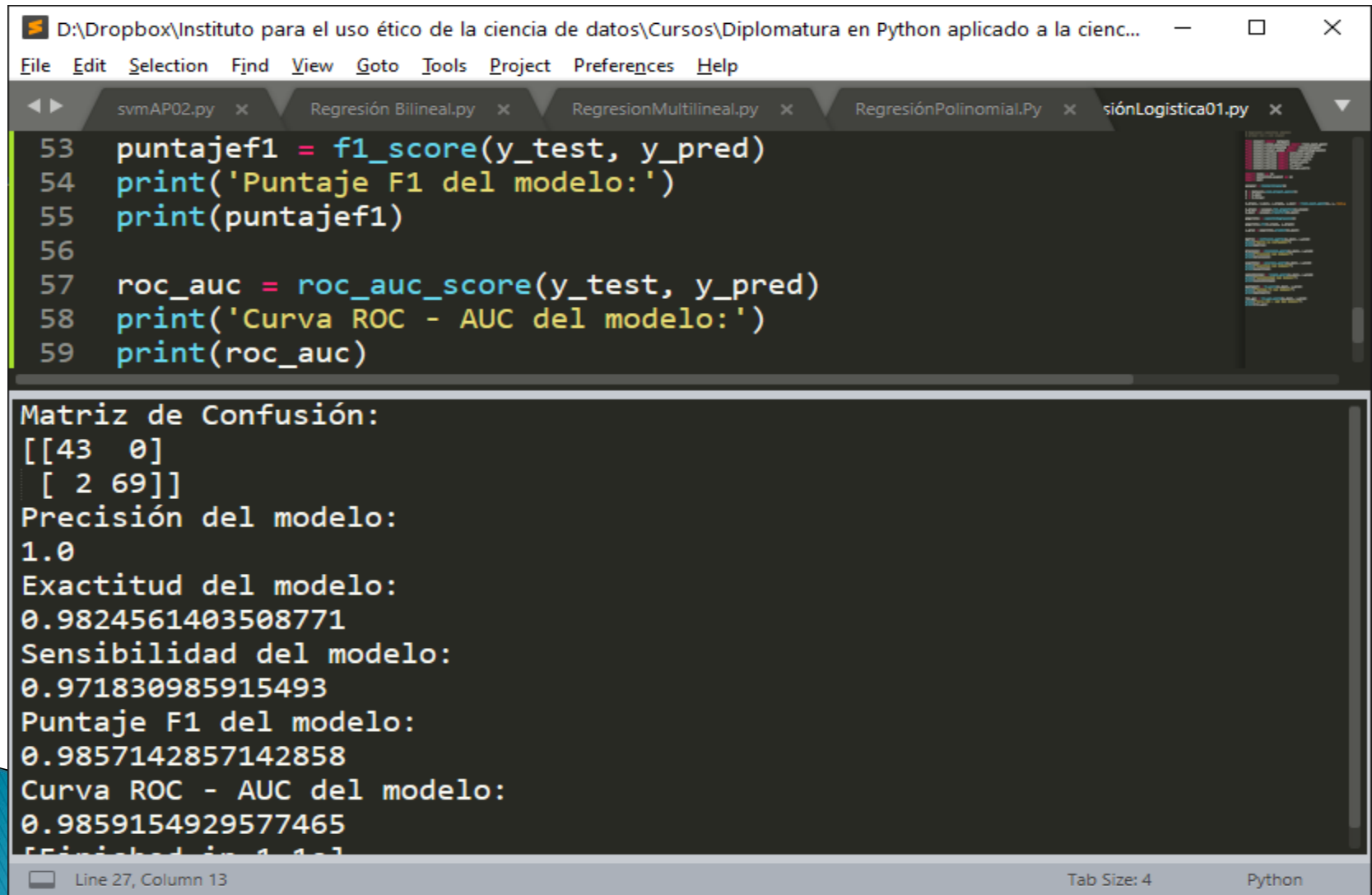


The image shows a screenshot of a Python IDE window. The title bar indicates the file path: D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia... The menu bar includes File, Edit, Selection, Find, View, Goto, Tools, Project, Preferences, and Help. The tab bar shows several open files: svmAP02.py, Regresión Bilineal.py, RegresiónMultilineal.py, RegresiónPolinomial.py, and siónLogistica01.py. The main editor area displays Python code for evaluating a logistic regression model. The code includes imports for confusion_matrix, precision_score, accuracy_score, and recall_score from sklearn.metrics. It then calculates and prints the confusion matrix, precision, accuracy, and recall for the model's predictions on a test set.

```
37 matriz = confusion_matrix(y_test, y_pred)
38 print('Matriz de Confusión:')
39 print(matriz)
40
41 precision = precision_score(y_test, y_pred)
42 print('Precisión del modelo:')
43 print(precision)
44
45 exactitud = accuracy_score(y_test, y_pred)
46 print('Exactitud del modelo:')
47 print(exactitud)
48
49 sensibilidad = recall_score(y_test, y_pred)
50 print('Sensibilidad del modelo:')
51 print(sensibilidad)
--
```

The status bar at the bottom shows 'Line 27, Column 13', 'Tab Size: 4', and 'Python'. A small globe icon is visible in the bottom right corner.

Regresión logística



The screenshot shows a Python IDE window with the title bar "D:\Dropbox\Instituto para el uso ético de la ciencia de datos\Cursos\Diplomatura en Python aplicado a la ciencia...". The menu bar includes File, Edit, Selection, Find, View, Goto, Tools, Project, Preferences, and Help. The tab bar shows several files: svmAP02.py, Regresión Bilineal.py, RegresiónMultilineal.py, RegresiónPolinomial.py, and siónLogistica01.py. The main editor displays the following Python code:

```
53 puntajef1 = f1_score(y_test, y_pred)
54 print('Puntaje F1 del modelo:')
55 print(puntajef1)
56
57 roc_auc = roc_auc_score(y_test, y_pred)
58 print('Curva ROC - AUC del modelo:')
59 print(roc_auc)
```

Below the code, the output of the program is displayed in a black console window:

```
Matriz de Confusión:
[[43  0]
 [ 2 69]]
Precisión del modelo:
1.0
Exactitud del modelo:
0.9824561403508771
Sensibilidad del modelo:
0.971830985915493
Puntaje F1 del modelo:
0.9857142857142858
Curva ROC - AUC del modelo:
0.9859154929577465
Finished in 1.101s
```

The status bar at the bottom indicates "Line 27, Column 13", "Tab Size: 4", and "Python".



Muchas Gracias

