Actividad #4 (Ajuste de Datos)

Instructor: Carlos Lizárraga Celaya *Student: Antonio Cota Rodríguez*

Introducción

Ajuste de Datos

El ajuste de curvas consiste en encontrar una curva que contenga una serie de puntos y que posiblemente cumpla una serie de restricciones adicionales. Esta sección es una introducción tanto a la interpolación (cuando se espera un ajuste exacto a determinadas restricciones) y al ajuste de curvas o análisis de regresión (cuando se permite una aproximación).

El ajuste de curvas es un proceso mediante el cual, dado un conjunto de N pares de puntos (x_i, y_i) (siendo x la variable independente e y la dependiente), se determina una función matemática $f(x)$ de tal manera que la suma de los cuadrados de la diferencia entre la imagen real y la correspondiente obtenida mediante la funcin ajustada en cada punto sea mnima :

$$\epsilon = min\left(\sum_{i=1}^{N}(y_i - f(x_i))^2\right)$$

Generalmente, se escoge una función genérica $f(x)$ en función de uno o más parámetros y se ajusta el valor de estos parámetros de la manera que se minimice el error cuadrático, ϵ. La forma más típica de esta función ajustada es la de un polinomio de grado M ; obteniéndose para $M = 1$ un ajuste lineal (o regresión lineal),

$$f(x) = a_0 + a_1 x$$

para M = 2 un ajuste parablico,

$$f(x) = a_0 + a_1 x + a_2 x^2$$

y así sucesivamente.

Ajuste por mínimos cuadrados

Mínimos cuadrados es una técnica de análisis numérico enmarcada dentro de la optimización matemática, en la que, dados un conjunto de pares ordenados variable independiente, variable dependiente y una familia de funciones, se intenta encontrar la función continua, dentro de dicha familia, que mejor se aproxime a los datos (un "mejor ajuste"), de acuerdo con el criterio de mínimo error cuadrático.

Existen numerosas leyes físicas en las que se sabe de antemano que dos magnitudes x e y se relacionan a través de una ecuación lineal

$$y = \mathbf{a}x + \mathbf{b}$$

donde las constantes **b** (ordenada en el origen) y **a** (pendiente) dependen del tipo de sistema que se estudia y, a menudo, son los parámetros que se pretende encontrar. El método de mínimos cuadrados determina los valores de los parámetros **a** y **b** de la recta que mejor se ajusta a los datos experimentales. Sin detallar el procedimiento, se dará aquí simplemente el resultado :

$$\mathbf{a} = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2}$$

$$\mathbf{b} = \frac{(\sum y_i) - \mathbf{a}(\sum x_i)}{n}$$

donde n es el número de medidas y Σ representa la suma de todos los datos que se indican.

Ajuste exponencial

Cuando la curva de regresión de y sobre x es exponencial, es decir para cualquier x considerada, la media de la distribución está dada por la siguiente ecuación predictora :

$$y = Ab^x$$

Tomando el logaritmo en ambos miembros de la ecuación :

$$\log y = \log A + x \log b$$

Y se puede estimar ahora $\log A$ y $\log b$, y de ahí obtener A y b aplicando el métodos de mínimos cuadrados.

Donde las constantes A y b quedan fijadas al resolver simultáneamente las ecuaciones :

$$\sum \log y = N \sum \log A + \log b \sum x$$

$$\sum x \log y = \log A \sum x + \log_b \sum x^2$$

Programas

Temperatura de invierno en Nueva York (Ajuste lineal)

El siguiente programa se utilizó el ajuste lineal por el métodos de mínimos cuadrados para las temperaturas medias registradas durante el invierno en la ciudad de Nueva York desde el año 1900 hasta 1999. El código que es (en Python) :

```
import numpy as np
from scipy import optimize
import matplotlib.pyplot as plt


#Archivo de datos
data = np.loadtxt('NewYork.txt')

x=data[:,0].astype(np.int)
y=data[:,1].astype(np.float)

#Escala
xn=np.linspace(1900,2000,2)

#Modulo polinomial
m, c = np.polyfit(x, y, 1)
#Aproximacin de los coeficientes
yn = np.polyval([m, c], xn)

#Imprimir la grafica
plt.plot(x, y, 'or', label="Datos")
plt.plot(xn, yn, label="Ajuste")
plt.title('Temperatura en NewYork')
plt.legend()
plt.show()
```

La gráfica ya ajustada es la siguiente :

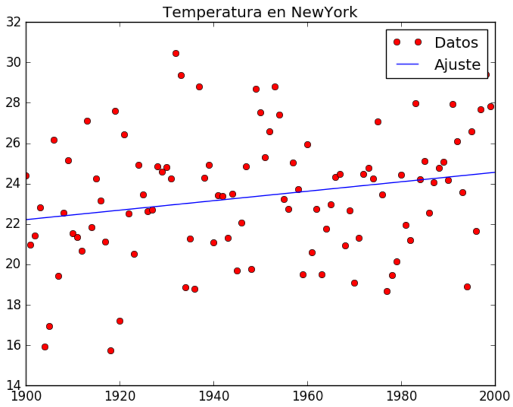

FIGURE 1 – Ajuste lineal

Presión atmosférica vs. altitud (Ajuste exponencial)

En esta sección se muestra un código para Python, en el cual se observa una relación entre la presión atmosférica y la altitud, estos datos obtenidos del artículo *Arguado E and Burt JE (1999), Understanding Weather and Climate ; Prentice Hall, Saddle River, New Jersey.*. El código es el siguiente :

```
import numpy as np
from scipy import optimize
import matplotlib.pyplot as plt

#Archivo de datos
data1=np.loadtxt('Presin_Altitud.txt')

x1=data1[:,0].astype(np.int)
y1=data1[:,1].astype(np.float)


#Definiendo la funcion
def f(x,u,v,w):
    return u*np.exp(-v*x) + w

#Opitimizar la curva
popt, pcov = optimize.curve_fit(f, x1, y1)



xm=np.linspace(-0,50,1000)

plt.plot(x1, y1, 'or', label="Datos")
plt.plot(xm,f(xm,*popt), label="Ajuste")

#Imprimir la grafica
plt.title('Presin-Altitud')
plt.legend()
plt.xlabel("Altitud")
plt.ylabel("Presin")
plt.show()
```

con su respectiva gráfica ajustada :

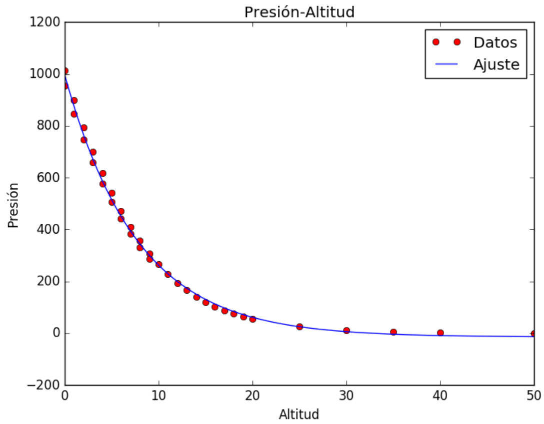

FIGURE 2 – Ajuste exponencial

Conclusión

En esta práctica aprendimos una potente herramienta computacional para el manejo y visualizacin de datos que nos será de gran ayuda durante nuestra vida acadmica como fuera de ella.