

GUÍA 4

Ajustes

Algoritmos

- I Programe el algoritmo que le permita resolver mínimos cuadrados para una recta, tome como entrada los datos x_i e y_i y devuelva los parámetros (a, b) de la recta $y = b + ax$.
- II Es posible generalizar la expresión de mínimos cuadrados para un polinomio de orden n puesto, que la primer fila tiene la suma sobre las potencias de x hasta el orden n igualado a la suma de los y . la segunda antes de realizar la suma multiplica todo por x , la tercera por x^2 y así hasta la última fila que multiplica todo por x^n . Dejando el sistema de la siguiente forma:

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 & \dots & \sum x_i^n \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \dots & \sum x_i^{n+1} \\ \vdots & & & & \vdots \\ \sum x_i^n & \sum x_i^{n+1} & \sum x_i^{n+2} & \dots & \sum x_i^{n+n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum y_i * x_i \\ \vdots \\ \sum y_i * x_i^n \end{bmatrix}$$

Cree una función que realice el mínimos cuadrados de orden n y devuelva el vector de coeficientes.

- III Cree una función que permita calcular el r^2 para un modelo polinomial de orden n , debe tomar como entrada los x , los y y los coeficientes a en un vector y devolver el valor de r^2 .
- IV Subir indiscriminadamente el orden del polinomio a ajustar consigue que se reduzca el error residual de ajuste. Pero es posible ver que esto no implica un mejor ajuste. Si se divide el set de datos en 2 partes y se usa una para realizar el ajuste y luego se calcula el r sobre los datos utilizados y sobre los datos que quedaron afuera puede verse un comportamiento peculiar.

Cree un algoritmo que permita separar los datos en un conjunto de entrenamiento o ajuste y otro conjunto de validación. debe tomar como entrada el conjunto total de x , el conjunto total de y y un porcentaje que será el de datos de entrenamiento. Como retorno de la función será $[x_{train}, y_{train}, x_{test}, y_{test}]$.

Ejercicio 1

Realice un ajuste de mínimos cuadrados con una recta al conjunto de puntos obtenidos de la función "e1Data.m" (llamela utilizando: $x, y = e1Data()$). Calcule el r^2 , grafique el conjunto de datos junto con el ajuste. Grafique el error de predicción de cada punto ($e_i = y_i - a_0 - a_1 x_i$).

Ejercicio 2

Utilice los datos de "e1Data.m" y realice ajustes con polinomios de grado 1, 2, 3, 4 y 5. Calcule en cada caso el r^2 . ¿Cuál elegiría para representar los datos?.

Ahora particione los datos en un conjunto de ajuste con el 70% de los datos, realice los mismos ajustes, calcule el r^2 sobre los datos que quedaron afuera. ¿Que sucede si compara los valores de los coeficientes de determinación?

Ejercicio 3

Se sabe que el conjunto de datos generado por "e3Data.m" ajusta bien bajo la fórmula $y = \frac{a}{1+b*x}$, ¿Qué transformación utilizaría para llevar el problema a una relación lineal? Aplíquela, encuentre el valor de los parámetros a , y b . Grafique el problema en el espacio transformado y en el espacio de origen.

Ejercicio 4

Para estimar el tamaño de una represa, es fundamental conocer el caudal historico del rio. Este dato muchas veces es imposible de conseguir. Los datos meteorológicos, sin embargo, suelen ser de facil acceso. Si se logra, entonces, determinar la relación entre caudal y precipitación, se puede realizar una estimación del caudal.

Los siguientes datos fueron obtenidos de los últimos 8 años:

Prec (cm/año)	99.3	116.8	94.2	126.8	139.6	104.3	108.4	88.6
Caudal (m3/s)	16.5	18.2	1.62	19.6	23.1	15.4	16.8	14.5

- Graficar Caudal vs Precipitación.
- Ajustar los puntos con una regresión lineal.
- Luego de ver el gráfico, usted supone que una de las mediciones esta mal tomada. Quite ese punto y vuelva a realizar el ajuste. ¿Ve alguna mejora?
- Utilizando los datos obtenidos del 2do ajuste, calcule el caudal para una precipitación de 115 cm/año.
- En un record histórico de precipitaciones, se registraron 230 cm/año. ¿Es posible estimar el caudal para esa valor de precipitaciones? Justifique.

Ejercicio 5

El siguiente modelo puede aplicarse a reacciones quimicas que ocurren en un biorreactor:

$$c = c_0 - \frac{1}{\sqrt{1 + 2kc_0^2t}}$$

donde c = concentración (mg/L), c_0 = concentración inicial (mg/L), k = tasa de reacción (L^2/mg^2d) y t = tiempo (d). Linearizar este modelo y usarlo para estimar k y c_0 , usando los siguientes datos:

tiempo	0	0.5	1	1.5	2	3	4	5
concentración	3.26	2.09	1.62	1.48	1.17	1.06	0.9	0.85

Table 1:

Ejercicio 6

En la función "e6Data.m" están los datos de un conjunto de puntos en el espacio (llamela utilizando $x, y, z = e6Data()$). Se sabe que este conjunto de puntos ajusta bien a un plano de la forma $z = a*x + b*y + c$. En clase no se vió el desarrollo para ajustes en varias dimensiones, pero están proporcionadas las herramientas para poder calcular los valores de los coeficientes. Para ello:

- Escriba el residual del ajuste.
- Derive respecto a los coeficientes para hallar los valores que lo minimizan.
- Construya y resuelva dicho sistema de ecuaciones para encontrar los valores de a, b y c .