

Práctica 5: Mínimos cuadrados

1. Determinar la recta que mejor aproxima, en el sentido de cuadrados mínimos, los siguientes conjuntos de datos

a)

x	1	2	3	4
y	4	4	3	1

b)

x	-1	0	1	3
y	0	1	3	7

2. Los siguientes datos corresponden a la población argentina (expresada en millones de habitantes):

Año	1950	1960	1970	1980	1991	2001	2010
Hab. (millones)	17	20.5	23.9	27.9	32.6	36.3	40.1

- a) Hallar una función f de la forma $f(x) = ax + b$ que mejor ajuste los datos en el sentido de mínimos cuadrados. Graficar en una misma figura los datos y la función f .
- b) Utilizando la función del ítem anterior ¿qué población se puede estimar que habrá en Argentina en el año 2025?
3. Implementar una función que reciba como input dos vectores de datos x e y y devuelva la recta de regresión lineal (como función de tipo `poly1d`).
4. a) Supongamos que se deja caer un objeto desde una altura de 200 m. Mientras cae, se toman las siguientes mediciones:

tiempo	0	1	2	4	6
altura	200	195	180	120	25

Hallar los coeficientes de la función $f(t) = at^2 + b$ que mejor aproxima los datos en el sentido de cuadrados mínimos.

- b) Sabiendo que la altura de dicho objeto, después de haber transcurrido un tiempo t , viene dada por $f(t) = 200 - \frac{1}{2}gt^2$, determinar el valor aproximado de g a partir de los resultados obtenidos.

5. Implementar una función que reciba una tabla de datos y nos devuelva el R^2 de aproximar esa lista por la recta de regresión lineal.
6. Calcular R^2 para la aproximación obtenida en los Ejercicios 1 y 2.
7. Implementar una función que reciba vectores de datos x, y y un vector de pesos positivos w y devuelva la función $f(x) = ax + b$ que minimiza el error

$$\sum_{0 \leq i \leq n} w_i (y_i - f(x_i))^2.$$

8. Desde 1981 hasta 1984 se estimó una población de conejos en una isla y se obtuvieron los datos

año	1981	1982	1983	1984
N	2960	4540	8080	17060

Se espera que los datos se ajusten a una función exponencial de la forma $N(t) = e^{at+b}$

- a) Usar el método de mínimos cuadrados para hacer este ajuste y calcular el error cometido.
 - b) Hacer ahora el ajuste utilizando el método de mínimos cuadrados generalizado y comparar.
 - c) Graficar simultáneamente los datos de la tabla y las funciones obtenidas.
9. Se quiere aproximar la siguiente tabla de datos:

1	2	3	6	7	8
0.268	0.218	0.185	0.129	0.118	0.109

con una función del tipo $f(x) = 1/(ax + b)$.

- a) Usar el método de mínimos cuadrados para hacer este ajuste y calcular el error cometido.
 - b) Hacer ahora el ajuste utilizando el método de mínimos cuadrados generalizado y comparar.
 - c) Graficar simultáneamente los datos de la tabla y las funciones obtenidas.
10. En cierta especie animal se estudia la relación entre el peso X (en kgs) y el volumen pulmonar Y (en litros), obteniéndose los datos

peso (kgs)	60	85	100	150	250
vol. pulmonar (l)	2.3	4	5	9	19.5

- a) Ajustar los datos a una función $Y = aX^b$ en el sentido de cuadrados mínimos. Calcular el error cometido.
- b) Ajustar los datos a una función $Y = aX^b$ en el sentido de cuadrados mínimos generalizados. Calcular el error cometido y comparar.
- c) Graficar simultáneamente los datos de la tabla y las funciones obtenidas.
11. Implementar un programa que reciba como input dos vectores x, y y $n \in \mathbb{N}$ y devuelva el polinomio de grado menor o igual a n que mejor ajusta en el sentido de mínimos cuadrados (como función de tipo `poly1d`).
12. Implementar una función que reciba como input dos vectores x e y y una lista de funciones y devuelva los coeficientes de la combinación lineal de las funciones que mejor ajuste los datos en el sentido de cuadrados mínimos.
13. Se quiere aproximar la siguiente tabla de datos:

1	2	4	5	7	8	11	12
1	1	2	2	3	3	4	4

con una función del tipo $f(x) = c_0 + c_1x + c_2\cos\left(\frac{\pi}{2}x\right) + c_3\sin\left(\frac{\pi}{2}x\right)$.

- a) Usar el método de mínimos cuadrados para hacer este ajuste.
- b) Graficar simultáneamente los datos de la tabla y la función obtenida.
14. El archivo `humedad.csv` contiene los datos (simulados) de porcentaje de humedad diaria de Buenos Aires en 2020. Como los datos de humedad muestran periodicidad anual, se propone un modelo
- $$f(t) = c_1 + c_2 \sin\left(t \frac{2\pi}{366}\right) + c_3 \sin\left(t \frac{4\pi}{366}\right).$$
- a) Usar el método de mínimos cuadrados para hacer este ajuste.
- b) Graficar simultáneamente los datos de la tabla y la función obtenida.
15. En el archivo `infantesConBajoPesoAlNacer.txt` se encuentran los datos correspondientes a mediciones de 100 niños nacidos con bajo peso en Boston (Labor and deliver characteristics and the risk of germinal matrix hemorrhage in low birth weight infants. Journal of child neurology, 6(1), 35-40, (1991))
- Llamamos

Y = perímetro cefálico del bebé al nacer, en centímetros (columna `headcirc`)

X_1 = edad gestacional del bebé al nacer, en semanas (columna `gestage`)

X_2 = peso al nacer del bebé, en gramos (columna `birthwt`)

Se quiere predecir el perímetro cefálico de un niño al nacer.

- a) Graficar X_1 vs Y e X_2 vs Y .
- b) Plantear un modelo de regresión lineal para predecir el perímetro cefálico del bebé en función de su edad gestacional. Calcular el error cometido.
- c) Plantear un modelo de regresión lineal múltiple para predecir el perímetro cefálico del bebé en función de su edad gestacional y de su peso al nacer. Calcular el error cometido y comparar.
- d) Si en el modelo obtenido en el ítem anterior mantenemos constante la edad gestacional, ¿cuántos centímetros de aumento en su perímetro cefálico, en promedio, se corresponde a cada incremento del peso en 10 gramos?