

## Regresión Lineal Simple

### Modo Directo

1. Dados los siguientes pares de datos  $(x, y)$  determinar la pendiente y ordenada al origen de la recta que hará el mejor ajuste lineal a dichos datos. (Rta.: La pendiente tendrá un valor aprox de 10.28, la ordenada de -0.93 y la función Costo un valor aprox. de 18.95; Esto puede variar si modifica la forma y el área donde busca los valores de  $m$  y  $b$ )  
Estamos suponiendo  $\hat{y} = mx + b$  será la recta de ajuste lineal.

x	y
2	18
3	33
4	37
5	54
6	59
7	71

*TIP: Puede hacer un chequeo, ya que son pocos datos, con Excel, utilizando las funciones de Excel  $SLOPE(ys, xs)$  e  $INTERCEPT(ys, xs)$  que calculan la pendiente  $m$ , y la ordenada al origen  $b$ , respectivamente, para ir chequeando su código esté obteniendo valores aprox. correctos.*

2. Pesos según Alturas. Dado el dataset **heights\_weights.csv** (un archivo de texto con valores separados por coma) que da el promedio de peso de un grupo de mujeres según su peso, en un muestreo de mujeres norteamericanas de 30 a 39 años, determinar la función de ajuste lineal simple para poder predecir según la altura de una mujer, el peso que tendrá. Mostrar los valores de la función Costo. Graficar los datos y el ajuste lineal. (la altura está en metros, el peso en kg.) Estime cuánto es el peso de una mujer de 1.30m, 1.60m y 2m.
3. Nota según cantidad de horas de estudio. Se cuenta con el dataset **notas\_horasestudio.csv** que contiene dos columnas horas, notas. Ese es el número de horas que un estudiante estudió y la nota que obtuvo al rendir el examen. Se desea aplicar regresión lineal simple para poder predecir que nota obtendrá un alumno según la cantidad de horas que estudió.

### Variado

4. Dados los siguientes pares de datos  $(x, y)$  determinar la pendiente y ordenada al origen de la recta que hará el mejor ajuste lineal a dichos datos. Estamos suponiendo  $\hat{y} = mx + b$  será la recta de ajuste lineal.

Resuelva este ejercicio de tres modos:

- a) Utilizando las fórmulas directas para calcular la pendiente y ordenada, para un ajuste lineal de una sola variable.
- b) Utilizando la función Costo y la técnica de fuerza bruta. (Elija una zona para buscar los valores de  $m$  y  $b$ , por ejemplo entre -10 y 0 para  $m$ , y 100 a 200 para  $b$ )

x	y
4.1	90
6	89
6	83
7.3	81
7.8	73
8.1	71
8.77	68
9.45	60
12.3	59
14.2	54
15	35
16	35
18.12	21

- c) Utilizando la función Costo y la técnica de Descenso por el Gradiente.
5. Dado el dataset adjunto con el práctico **fecha\_dolar\_bna.csv** de valores del dólar oficial del BNA, desde el año 2015 hasta el año 2023, determinar su ajuste lineal. Resolver utilizando Descenso por el Gradiente. Mostrar la gráfica de datos con el ajuste. Predecir cuánto saldrá el dólar en el mes Noviembre 2023, Marzo de 2024 y Abril 2024 (*Considere ingresar las fechas como str, del tipo '01/11/2023' y convertirlas a su equivalente valor numérico int.*  
*En la serie de datos en el archivo .csv puede modificar las fechas dentro de Excel, modificando el formato de celdas a tipo Número y las fechas quedarán convertidas automáticamente a su equivalente numérico, o bien puede hacer una asociación de valores, donde 1 represente la primer fecha del dataset, y así, x vaya incrementando, o bien, puede investigar la librería datetime de python. O bien, puede extraer el día, mes, y año, y sacar cuentas para obtener su número. Diseñe dos funciones auxiliares para el manejo de esta parte. Una que de la fecha en str, obtenga su int asociado. Otra que dado un int, retorne su fecha asociada. Requiere maña.)(. ([Link oficial dataset BNA](#))*

## Regresión Lineal Múltiple - Descenso por el Gradiente

6. Para los siguientes datos, determine el mejor ajuste lineal y el valor de la función Costo, en 100, 1000, 100000 iteracciones. Grafique la función Costo respecto al número de iteracciones. En este caso, debería encontrar que  $b = 1$  y los coeficientes serán  $m_1 = 2, m_2 = 3, m_3 = 4$  y la función Costo tenderá a cero. (*tenemos una solución exacta porque hemos partido de datos bien elegidos a propósito*)

```
xss = [[1,1,1], [1,3,4], [3,4,5], [2,2,2], [3,3,3], [0,1,1]]  
ys = [10, 28, 39, 19, 28, 8]
```

7. Bob ha puesto una fábrica de teléfonos móviles. Necesita encontrar algún criterio para determinar el rango de precios de sus celulares. Ha juntado data de varias compañías importantes, y el rango de precios (0 = *muybajo*, 1 = *bajo*, 2 = *medio*, 3 = *alto*) Se

debe predecir el rango de precios de un teléfono, no el precio exacto. Seleccionar al menos 4 características de todas las disponibles en el dataset *celulares.train.csv* y realizar el ajuste múltiple, por Descenso por el Gradiente. Luego, predecir para los siguientes casos, qué rango de precio tendrían. *(Tomar solo las columnas que haya seleccionado en su ajuste)*

bateria	bluetooth	clock	dual_sim	camara_frontal	4G	memoria_interna
1043	1	1.8	1	14	0	5
841	1	0.5	1	4	1	61
1807	1	2.8	0	1	0	27
1546	0	0.5	1	18	1	25
1434	0	1.4	0	11	1	49
1464	1	2.9	1	5	1	50
1718	0	2.4	0	1	0	47
833	0	2.4	1	0	0	62

peso	nucleos	pc	altura_pixeles	ancho_pixeles	RAM
193	3	16	226	1412	3476
191	5	12	746	857	3895
186	3	4	1270	1366	2396
96	8	20	295	1752	3893
108	6	18	749	810	1773
198	8	9	569	939	3506
156	2	3	1283	1374	3873
111	1	2	1312	1880	1495