Departamento de Ciencia de la Computación e Inteligencia Artificial

Alumno:	
Grupo teoría:	
DNI:	

Ejercicio RESUELTO teoría Bloque Resolución Ecuaciones. Matemáticas II, 2019/20 GRUPO ?

Instrucciones generales:

Todos los ejercicios deben estar bien explicados. Entregar todas las hojas firmadas y con DNI.

EJERCICIO 2. Se tiene la recta que viene dada por los puntos $(x_1, y_1) = (2,2)$ y $(x_2, y_2) = (3,7)$ se pide

a) Calculad la ecuación de la recta que pasa por esos dos puntos en forma $y = mx + b$

b) Suponed que incluimos un tercer punto $(x_3, y_3) = (2,4)$. Determinad, mediante descenso por gradiente, si la recta que minimiza el error cuadrático se parece a la que se obtiene en el apartado a). Para ello hay que resolver (aproximadamente) el siguiente sistema:

$$\begin{aligned} mx_1 + b &= y_1 \\ mx_2 + b &= y_2 \\ mx_3 + b &= y_3 \end{aligned}$$

Usad como punto inicial $(m_0 = m, b_0 = b)$ es decir los parámetros obtenidos en a), usad la constante de ponderación $\gamma = 0.1$ y retened 4 decimales en cada iteración.
REALIZAR 4 ITERACIONES Y MEDID EL ERROR ENTRE UNA ITERACIÓN Y LA ANTERIOR.

c) Repetid el apartado b) suponiendo $(x_3, y_3) = (4,0)$. Explicad los cambios observados.

a) Dados $(x_1, y_1) = (1,2)$ y $(x_2, y_2) = (3,7)$ tenemos que la recta (m,b) sale de:

$$\frac{(y_2 - y_1)}{(x_2 - x_1)} = m; \Rightarrow m = \frac{5}{1} = 5$$

$$(y - y_1) = m(x - x_1) \ \Rightarrow (y - 2) = m(x - 2) \ \Rightarrow y = 5x - 8 \ \Rightarrow m = 5, b = -8$$

b) El tercer punto $(x_3, y_3) = (2,4)$ no pertenece claramente a la recta ya que $4 \neq 5 \cdot 2 - 8 = 2$. Como consecuencia, el sistema , aunque sea lineal

$$\begin{aligned} m \cdot 2 + b &= 2 \\ m \cdot 3 + b &= 7 \\ m \cdot 2 + b &= 4 \end{aligned}$$

Es incompatible (inconsistente) por lo que solo se puede resolver de forma aproximada. Para ello usamos la **función de error cuadrática** que vimos en la clase de descenso por gradiente:

$$E(m, b) = \frac{1}{2} \sum_{i=1}^{n} (mx_i + b - y_i)^2$$

Es la suma de los cuadrados de las ecuaciones (una por cada punto o dato en este caso) del sistema homogéneo equivalente (dejando ceros en la parte derecha). En detalle:

$$E(m,b)=\frac{1}{2}[(mx_1+b-y_1)^2+(mx_2+b-y_2)^2+(mx_3+b-y_3)^2]$$

Sustituyendo:

$$E(m,b)=\frac{1}{2}[(2m+b-2)^2+(3m+b-7)^2+(2m+b-4)^2]$$

1. **Calcular el gradiente de la función de error:** $\nabla E(m,b)=[\frac{\partial E}{\partial m}\ \ \frac{\partial E}{\partial b}]^T$ que viene dado por

$$\frac{\partial E}{\partial m}=(2m+b-2)2+(3m+b-7)3+(2m+b-4)2$$
$$\frac{\partial E}{\partial b}=(2m+b-2)1+(3m+b-7)1+(2m+b-4)1$$

Recomendamos reagrupar de la siguiente manera:

$$\begin{aligned}\frac{\partial E}{\partial m} &= (2m+b-2)2+(3m+b-7)3+(2m+b-4)2\\ &= (4m+2b-4)+(9m+3b-21)+(4m+2b-8)\\ &= 17m+7b-33\end{aligned}$$

$$\begin{aligned}\frac{\partial E}{\partial m} &= (2m+b-2)1+(3m+b-7)1+(2m+b-4)1\\ &= (2m+b-2)+(3m+b-7)+(2m+b-4)\\ &= 7m+3b-13\end{aligned}$$

Para esta función de error tenemos una **fórmula general** en función del número de puntos:

$$\frac{\partial E}{\partial m}=m\sum_{i=1}^{n}x_i^2+b\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}x_iy_i$$
$$\frac{\partial E}{\partial b}=m\sum_{i=1}^{n}x_i+n\cdot b-\sum_{i=1}^{n}y_i$$

Así pues el gradiente para este apartado viene dado por

$$\frac{\partial E}{\partial m}=17m+7b-33$$
$$\frac{\partial E}{\partial b}=7m+3b-13$$

2. **Hacer $t=0$ e iterar en la dirección opuesta al gradiente** según la ecuación:

$$\begin{bmatrix}m_{t+1}\\ b_{t+1}\end{bmatrix}=\begin{bmatrix}m_t\\ b_t\end{bmatrix}-\gamma\nabla E(m_t,b_t)$$

Que para nuestro problema queda así:

$$\begin{bmatrix}m_{t+1}\\ b_{t+1}\end{bmatrix}=\begin{bmatrix}m_t\\ b_t\end{bmatrix}-\gamma\begin{bmatrix}17m_t+7b_t-33\\ 7m_t+3b_t-13\end{bmatrix}$$

Suponemos $m_0 = 5, b_0 = -8$ **e iteramos para** $t = 1$

$$\begin{bmatrix} m_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} 5 \\ -8 \end{bmatrix} - 0.1 \begin{bmatrix} 17 \cdot 5 + 7 \cdot (-8) - 33 \\ 7 \cdot 5 + 3 \cdot (-8) - 13 \end{bmatrix} = \begin{bmatrix} 5 \\ -8 \end{bmatrix} - 0.1 \begin{bmatrix} -4 \\ -2 \end{bmatrix} = \begin{bmatrix} 5.4 \\ -7.8 \end{bmatrix}$$

Vemos que el vector gradiente tiene aún un módulo mucho mayor que cero.

Iteramos para $t = 2$

$$\begin{bmatrix} m_2 \\ b_2 \end{bmatrix} = \begin{bmatrix} 5.4 \\ -7.8 \end{bmatrix} - 0.1 \begin{bmatrix} 17 \cdot (5.4) + 7 \cdot (-7.8) - 33 \\ 7 \cdot (5.4) + 3 \cdot (-7.8) - 13 \end{bmatrix} = \begin{bmatrix} 5.4 \\ -7.8 \end{bmatrix} - 0.1 \begin{bmatrix} -3.92 \\ -1.96 \end{bmatrix} = \begin{bmatrix} 4.98 \\ -7.94 \end{bmatrix}$$

Vemos que el resultado se parece mucho al de la iteración anterior. No obstante seguimos iterando.

Iteramos para $t = 3$

$$\begin{bmatrix} m_3 \\ b_3 \end{bmatrix} = \begin{bmatrix} 4.98 \\ -7.94 \end{bmatrix} - 0.1\, \nabla E(4.98, -7.94) = \begin{bmatrix} 4.98 \\ -7.94 \end{bmatrix} - 0.1 \begin{bmatrix} -3.92 \\ -1.96 \end{bmatrix} = \begin{bmatrix} 5.372 \\ -7.744 \end{bmatrix}$$

Iteramos para $t = 4$

$$\begin{bmatrix} m_4 \\ b_4 \end{bmatrix} = \begin{bmatrix} 5.372 \\ -7.744 \end{bmatrix} - 0.1\, \nabla E(5.372, -7.744) = \begin{bmatrix} 5.372 \\ -7.744 \end{bmatrix} - 0.1 \begin{bmatrix} 4.116 \\ 1.372 \end{bmatrix} = \begin{bmatrix} 4.9604 \\ -7.8812 \end{bmatrix}$$

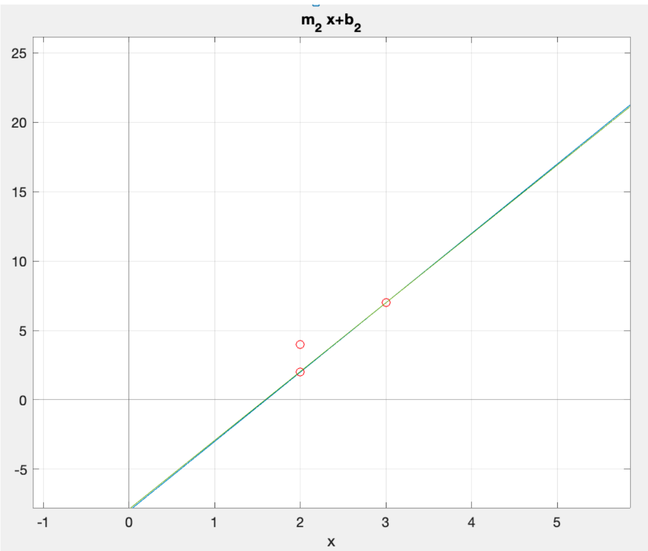
Después de la cuarta iteración vemos que el valor de $\sqrt{(m_{t+1} - m_t)^2 + (b_{t+1} - b_t)^2}$ (distancia entre dos soluciones sucesivas) es de 0.43 con lo cual podríamos parar aquí asumiendo un épsilon o tope de error de $\epsilon = 0.5$. La **solución** sería [4.9604 -7.8812], muy parecida a la recta inicial. Esto quiere decir que la recta que se ajusta a los tres puntos es muy similar a la que pasa por los dos primeros

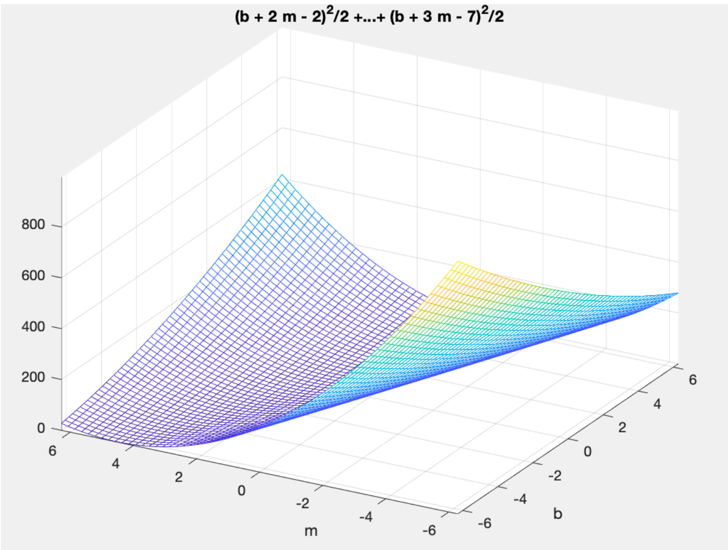
3. **EN FORMA DE TABLA**

Una forma resumida de poner estos datos es cumplimentar una tabla con tres columnas. En la primera ponemos el m,b de la iteración anterior, en la segunda el gradiente Grad(m,b) y en la tercera el nuevo m’,b’ tras multiplicar el Grad(m,b) por gamma y restar el resultado a la primera columna:

	m,b	**Grad(m,b)**	**m’,b’**
1	(5,-8)	(-4,2)	(5.4,-7.8)
2	(5.4,-7.8)	(-3.92,-1.96)	(4.98,-7.94)
3	(4.98,-7.94)	(-3.92,-1.96)	(5.372,-7.744)
4	(5.372,-7.744)	(4.116,1.372)	**(4.9604,-7.8812)**

4. **INTERPRETACIÓN GRÁFICA**
UTILIZANDO MATLAB podemos visualizar la recta solución con respecto a la recta original. Se superponen. La función de error, la podemos visualizar al lado:





La función de error para este apartado parece tener una zona plana. Si calculamos H la matriz hessiana 2x2 (se podría preguntar en el exámen) veríamos que su determinante es positivo y la componente H(1,1) >0. O sea que tenemos un mínimo. Concretamente la solución óptima es **m=4, b=-5.**
PODRÍAMOS HABER OBTENIDO ESTA MISMA SOLUCIÓN RESOLVIENDO EL SISTEMA:

$$\frac{\partial E}{\partial m} = 17m + 7b - 33 = 0$$
$$\frac{\partial E}{\partial b} = 7m + 3b - 13 = 0$$

b) El tercer punto $(x_3, y_3) = (4,0)$ no pertenece claramente a la recta ya que $0 \neq 5 \cdot 4 - 8 = 12$. En consecuencia, el sistema , aunque sea lineal

$$m \cdot 2 + b = 2$$
$$m \cdot 3 + b = 7$$
$$m \cdot 3 + b = 1$$

Es de nuevo incosistente. Nos piden aplicar descenso por gradiente para hallar m y b de la recta que pasa aproximadamente por esos 3 puntos. De nuevo la función de error es:

$$E(m, b) = \frac{1}{2}[(mx_1 + b - y_1)^2 + (mx_2 + b - y_2)^2 + (mx_3 + b - y_3)^2]$$

1. **Calcular el gradiente de la función de error:**

En general podemos aplicar las siguientes expresiones para $\nabla E(m, b) = [\frac{\partial E}{\partial m} \ \ \frac{\partial E}{\partial b}]^T$

$$\frac{\partial E}{\partial m} = m\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i$$
$$\frac{\partial E}{\partial b} = m\sum_{i=1}^{n} x_i + n \cdot b - \sum_{i=1}^{n} y_i$$

O sea

$$\frac{\partial E}{\partial m} = m(2^2 + 3^2 + 4^2) + b(2 + 7 + 0) - (2 \cdot 2 + 3 \cdot 7 + 0)$$
$$\frac{\partial E}{\partial b} = m(2 + 3 + 4) + 3 \cdot b - (2 + 7 + 0)$$

Así pues el gradiente para este apartado viene dado por

$$\frac{\partial E}{\partial m} = 29m + 9b - 25$$
$$\frac{\partial E}{\partial b} = 9m + 3b - 9$$

2. **Hacer $t = 0$ e iterar en la dirección opuesta al gradiente** según la ecuación**:**

$$\begin{bmatrix} m_{t+1} \\ b_{t+1} \end{bmatrix} = \begin{bmatrix} m_t \\ b_t \end{bmatrix} - \gamma \nabla E(m_t, b_t) = \begin{bmatrix} m_t \\ b_t \end{bmatrix} - \gamma \begin{bmatrix} 29m_t + 9b_t - 27 \\ 9m_t + 3b_t - 9 \end{bmatrix}$$

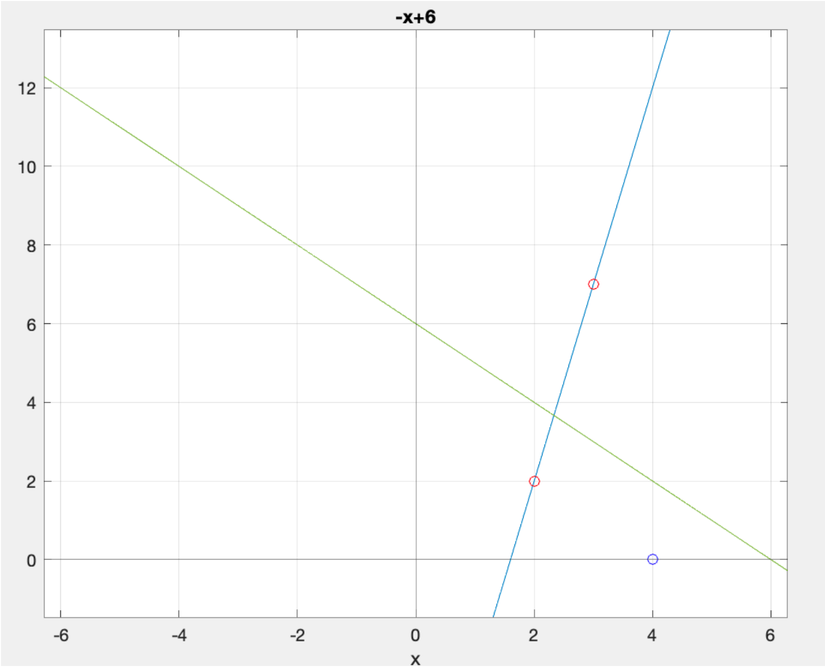
Partimos de $(m_0 = 5\ , b_0 = -8)$ así que usando la tabla tememos:

	m,b	**Grad(m,b)**	**m',b'**
1	(5,-8)	(48,12)	(0.2,-9.1999)
2	(0.2,-9.1999)	(-1.02,-0.348)	(10.4,-5.72)
3	(10.4,-5.72)	(225.12,67.44)	(-12.1120, -12.464)
4	(-12.1120, -12.464)	(-488.424,-155.4)	**(36.7304, 3.076)**

EXPLICACIÓN:
Claramente el método diverge, ¿por qué? La función de coste es similar a la del apartado a), es convexa y tiene un mínimo en **m=-1, b=6,** la solución del sistema $\nabla E(m, b) = [0\ 0]^T$**.**
Veamos qué ocurre:

* La solución óptima está lejos del punto de inicialización, pero aunque el error es convexo debería converger.
* Lo que ocurre es que la constante $\gamma = 0.1$ es demasiado grande y hace oscilar al método.
* De todas formas, ni siquiera haciendo $\gamma = 0.01$ converge. Tarda 12 iteraciones en llegar a **m=-0.4999** y **b=-0.4999.**
* En estos casos **(no se os pide nunca hacer muchas iteraciones y siempre podéis resolver el sistema de gradiente=0 que es 2x2 y os dice el óptimo si existe)**

CONVIENE REPRESENTAR GRÁFICAMENTE LOS DATOS Y VER QUÉ OCURRE: (página siguiente)



En azul la recta definida por $(x_1, y_1) = (2{,}2)$ y $(x_2, y_2) = (3{,}7)$. Al añadir el punto $(x_3, y_3) = (4{,}0)$, y partir de la recta $y = 5x - 8$, la forzamos a disminuir su pendiente (iteración 2 de la tabla). Sin embargo lo que ocurre es que al hacerlo, desajustamos los dos primeros puntos y eso nos lleva a hacer la pendiente casi vertical (iteración 3). Finalmente, en la iteración 3 la pendiente pasa a ser negativa, con valor m=-12.1120 pero aún le queda mucho para llegar al óptimo m=-1, b=6 es decir a la recta $y = -x + 6$. Puesto que b=-12.464 también es negativa. En resumen, hemos acabado en una recta en el tercer cuadrante (x e y negativos). Tan abajo que no la podemos dibujar. COMO PUEDE VERSE, DIBUJAR EL PROBLEMA AYUDA A INTERPRETAR LA SOLUCIÓN.