```
Alumno:
    Christian Hernandez
Docente:
        Ing. Vladimir Robles
Materia:
       Inteligencia Artificial II.
Tema
        GradientDescent-1
Ciclo
            10no Ciclo

    Cuenca Noviembre 2020
```


Descenso por gradiente definicion

El método del descenso del gradiente es un algoritmo de optimización que permite converger hacia el valor mínimo de una función mediante un proceso iterativo. En aprendizaje automático básicamente se utiliza para minimizar una función que mide el error de predicción del modelo en el conjunto de dato

Objetivos:

- Familiarizarse con los principales aspectos y etapas del método de descenso por gradiente.
- Conocer y aplicar el método de descenso por gradiente para minimizar funciones sencillas.
- Conocer cómo aplicar el proceso de descenso por gradiente de forma automatizada con soporte de **sympy** y **Jupyter Notebook*.

Criterios de evaluación:

Los criterios o rúbrica de evaluación del Boletín de Prácticas 1 son los siguientes:

1. Adecuada complejidad de la función seleccionada para el proceso de minimización (procurar no usar funciones cóncavas).
2. Correcta explicación y detalle de cada paso ejecutado con el método de descenso por gradiente.
3. Verificación de la solución encontrada.
4. Elementos extra: incluye recta (flecha) con la dirección del gradiente, animaciones, etc.

Prerrequisitos:

A fin de poder realizar esta práctica, deberá contar con los siguientes prerrequisitos:

1. Haber leído de forma completa el presente cuaderno.
2. Tener instalados los siguientes paquetes en su computador:
   A. Python 2.7+ (de preferencia 3.6+)
   B. Sympy
   C. matplotlib

Práctica 1:

Seleccionar una función matemática para realizar el proceso de minimización.

Ejecutar al menos 3 pasos del método de descenso por gradiente.

Diseñar y desarrollar un cuaderno en Jupyter Notebook donde se realicen todos los pasos correspondientes al proceso de minimización de la función a través del método de descenso por gradiente.

El cuaderno deberá incluir los siguientes puntos: Gráfica de la función y los puntos que se obtienen a medida que se ejecutan los pasos de cálculo (hasta k=3).

Aplicación de las funciones de derivación y evaluación de forma similar a la que se ha detallado en el presente cuaderno.

Incluir un acápite sobre las funciones cóncavas y los puntos estacionarios (incluir gráficos).

Emplear las funcionalidades que proveen los paquetes matplotlib y sympy.

Ejercicio a realizar

Funcion a minimizar

$\frac{1}{5.3}x^{3}-3x+x^{2}-2.37$

El valor de punto de partida será $x^{(0)}=5.3$

Graficamos la función que se minimizará y el punto de partida:

Importamos la liberias

In [1]:

```
import matplotlib.pyplot as pp
import numpy as np
```

Funcion

In [2]:

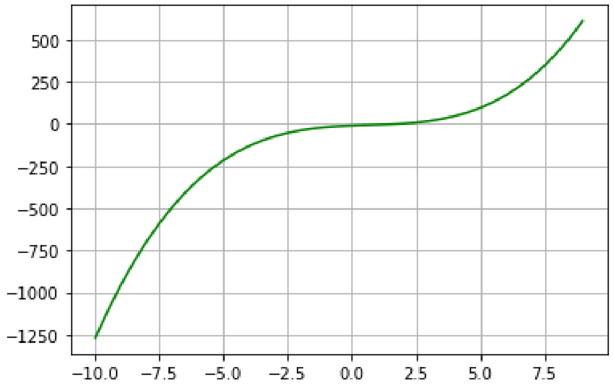
```
def fx(x):
    return np.power(x,3)-2*np.power(x,2)+6*x-10

x=np.arange(-10,9.5,0.5)
y=fx(x)
```

Grafica

In [3]:

```
pp.plot(x,y,c='green')
pp.grid(True)
pp.show()
a = 23
```



Funcion

In [4]:

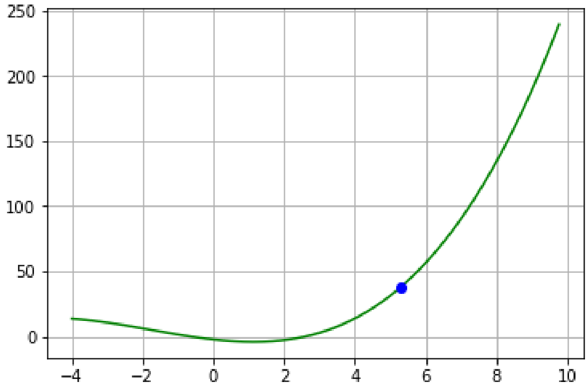
```
def fx2(x2):
    return (1/5.3)*np.power(x2,3)-3*x2+np.power(x2,2)-2.37

x2=np.arange(-4,10,0.27)
y2=fx2(x2)
```

Grafica

In [5]:

```
pp.plot(x2,y2,c='green')
pp.plot(5.3,fx2(5.3),'bo')
pp.grid(True)
pp.show()
```



Derivacion

- Realizamos derivada de la función y evaluamos su valor en el punto $x^{(0)}$:

Importamos la liberias

In [6]:

```
from sympy import Symbol, Function, diff, solve
```

In [7]:

```
x=Symbol('x')
f=Function('f')(x)
x0=5.3

fx=(1/5.3)*x**3-3*x+x**2-2.37

r=diff(fx)
```

Resultado

In [8]:

```
print("Derivada de la funcion: ",r)
print('****************************************************************')
print("Valor en x^(0):",r.subs(x,x0).evalf())
print('********************************')
```

```
Derivada de la funcion:  0.566037735849057*x**2 + 2*x - 3
*********************************************************
Valor en x^(0): 23.5000000000000
********************************
```

Lo que nos da como resultado lo siguiente:

$f^{'}(x) = 0.57x^{2}+2x-3$

$\nabla f(x^{(0)}) = 23.50$

- A fin de de encontrar el valor de t^{*}_{0} requerido para calcular el siguiente punto $x^{(1)}$, calculamos el valor de la siguiente función:

$\theta(t) = f(x^{(0)})-t \nabla f(x^{(0)})$.

$\theta^{'}(t)=-\nabla f\left(x^{(0)} - t \nabla f(x^{(0)} \right) \nabla f(x^{(0)})$

$=-\nabla f\left(\mathbf(5.3)-t \nabla f(x^{0}) \right) \nabla f(x^{(0)})$

$=-\nabla f\left(\mathbf(5.3)-23.50\cdot t \right) 23.50$

En este punto, sustituimos $(5.3-23.50\cdot t)$ por x en la función a minimizar que derivamos previamente $f^{'}(x) = 0.57x^{2}+2x-3$:

$=-\left(0.57\cdot(5.3-23.5t)^{2}+2 \cdot (5.3-23.5t)-3 \right)\cdot 23.50$

Aplicamos factoreo a la ecuación y calculamos las raíces

Ingresamos la funcion en sympy y buscamos las raices:

In [9]:

```
fp=-((0.57*(5.3-23.5*x)**2+2*(5.3-23.5*x)-3))
print(fp.expand())

print("Raices de la ecuacion:" ,solve(fp))
```

```
-314.7825*x**2 + 188.987*x - 23.6113
Raices de la ecuacion: [0.177289454026470, 0.423083819583086]
```

No realizamos la multiplicación por 23.5 ya que luego se igualará a 0, con ello tenemos lo siguiente:

$=-\left(-314.78x^{2}+188.99x-23.6113 \right)\cdot 23.5$

Y las raíces de la ecuación representarán 2 posibles valores para t^{*}_{0}:

$\left[0.18, 0.42 \right]$

Con ello, calculamos los dos posibles puntos $x^{(1)}$ y determinamos con cuál de los 2 se minimiza de mejor forma la función:
$ \begin{split} x^{(1)}=\begin{cases} 5.3-0.177\cdot 23.50~ = & \mathbf{1.14}\\ 5.3-0.423\cdot 23.50~ = & \mathbf{-4.64} \end{cases} \end{split} $

De igual forma, evaluamos la función original con los dos nuevos puntos $x^{(1)}$ y en base a ello determinamos cuál es el que minimiza de mejor manera su valor:

In [10]:

```
v1=fx.subs(x,1.14).evalf()
v2=fx.subs(x,-4.64).evalf()

print("Valores con de la funcion: [%d, %d]" % (v1,v2))
```

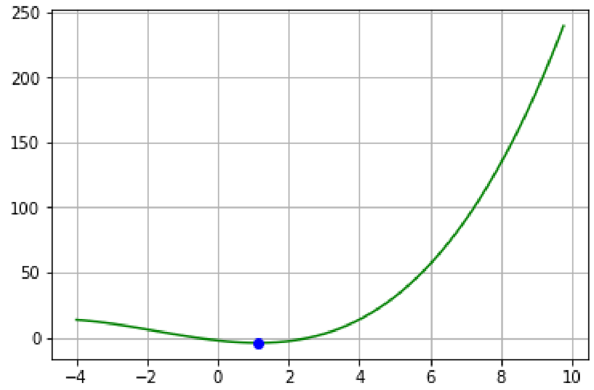
```
Valores con de la funcion: [-4, 14]
```

$ \begin{split} f{(x)}=\begin{cases} f(1.14)~ = & \mathbf{-4}\\ f(-4.6405)~ = & \mathbf{14} \end{cases} \end{split} $
Con ello, determinamos que la mejor alternativa es la de la segunda raíz para calcular el punto $x^{(1)}=1.14$ y hacemos $k \leftarrow k+1$.

Grafica

In [11]:

```
pp.plot(x2,y2,'r',color='green')
pp.plot(1.14,fx2(1.14),'bo')
pp.grid(True)
pp.show()
```



Derivada

In [12]:

```
x1 = 1.14
print("Derivada de la funcion: ",r)
print("Valor en x^(1):",r.subs(x,x1).evalf())
```

```
Derivada de la funcion:  0.566037735849057*x**2 + 2*x - 3
Valor en x^(1): 0.0156226415094338
```

Segunda Interaccion

$\nabla f(x^{(1)}) = 0.016$

$\theta(t) = f(x^{(1)})-t \nabla f(x^{(1)})$.

$\theta^{'}(t)=-\nabla f\left(x^{(1)} - t \nabla f(x^{(1)} \right) \nabla f(x^{(1)})$

$=-\nabla f\left(\mathbf(1.14)-t \nabla f(x^{1}) \right) \nabla f(x^{(1)})$

$=-\nabla f\left(\mathbf(1.14)-0.016\cdot t \right) 0.016$

En este punto, sustituimos $(1.14-0.016\cdot t)$ por x en la función a minimizar que derivamos previamente $f^{'}(x) = 0.57x^{2}+2x-3$:

$=-\left(0.57\cdot(1.14-0.016t)^{2}+2 \cdot (1.14-0.016t)-3 \right)\cdot 0.016$

Aplicamos factoreo a la ecuación y calculamos las raíces:

Ingresamos la funcion en sympy y buscamos las raices:

In [13]:

```
-0.00014592*x**2 + 0.0527936*x - 0.020771999999997
```

```
Raices de la ecuacion: [0.393885601377983, 361.404360012657]
```

No realizamos la multiplicación por 0.016 ya que luego se igualará a 0, con ello tenemos lo siguiente:

$=-\left(-0.00014x^{2}+0.052x-0.020 \right)\cdot 0.016$

Y las raíces de la ecuación representarán 2 posibles valores para t^{*}_{1}:

$\left[0.393, 361.40 \right]$

Con ello, calculamos los dos posibles puntos $x^{(1)}$ y determinamos con cuál de los 2 se minimiza de mejor forma la función: $ \begin{split} x^{(1)}=\begin{cases} 1.14-0.393\cdot 0.016~ = & \mathbf{1.133}\\ 5.3-361.40\cdot 0.016~ = & \mathbf{-4.6424} \end{cases} \end{split} $

De igual forma, evaluamos la función original con los dos nuevos puntos $x^{(1)}$ y en base a ello determinamos cuál es el que minimiza de mejor manera su valor:

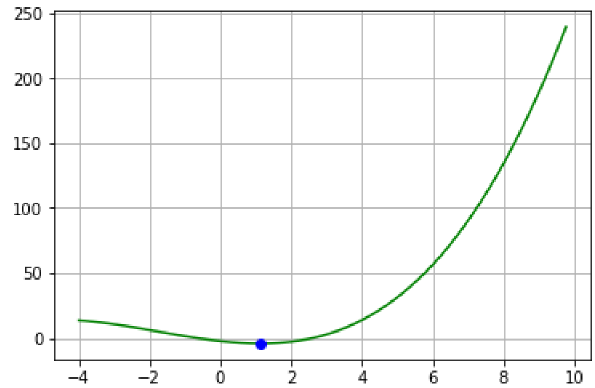
Funcion

In [14]:

```
Valores con de la funcion: [-4, 14]
```

Grafica

In [15]:



Derivada

In [16]:

```
Derivada de la funcion:  0.566037735849057*x**2 + 2*x - 3
Valor en x^(1): -0.00738358490566016
```

Tercera Interaccion

$\nabla f(x^{(2)}) = 0.0074$

$\theta(t) = f(x^{(2)})-t \nabla f(x^{(2)})$.

$\theta^{'}(t)=-\nabla f\left(x^{(2)} - t \nabla f(x^{(2)} \right) \nabla f(x^{(2)})$

$=-\nabla f\left(\mathbf(1.133)-t \nabla f(x^{1}) \right) \nabla f(x^{(1)})$

$=-\nabla f\left(\mathbf(1.133)-0.0074\cdot t \right) 0.016$

_ En este punto, sustituimos $(1.133-0.0074\cdot t)$ por x en la función a minimizar que derivamos previamente $f^{'}(x) = 0.57x^{2}+2x-3$:

$=-\left(0.57\cdot(1.133-0.0074t)^{2}+2 \cdot (1.133-0.0074t)-3 \right)\cdot 0.0074$

Aplicamos factoreo a la ecuación y calculamos las raíces:

Ingresamos la funcion en sympy y buscamos las raices:

In [17]:

```
-3.12132e-5*x**2 + 0.024357988*x + 0.00229727000000002
Raices de la ecuacion: [-0.0943014024259826, 780.468886513853]
```

No realizamos la multiplicación por 0.0074 ya que luego se igualará a 0, con ello tenemos lo siguiente:

$=-\left(-0.000003121x^{2}+0.02435x-0.0022 \right)\cdot 0.0074$

Y las raíces de la ecuación representarán 2 posibles valores para t^{*}_{1}:

$\left[0.0943, 780.47 \right]$

Con ello, calculamos los dos posibles puntos $x^{(1)}$ y determinamos con cuál de los 2 se minimiza de mejor forma la función: $ \begin{split} x^{(1)}=\begin{cases} 1.133-(-0.094)\cdot 0.0074~ = & \mathbf{1.1336956}\\ 5.3-780.47\cdot 0.0074~ = & \mathbf{-4.642478} \end{cases} \end{split} $

De igual forma, evaluamos la función original con los dos nuevos puntos $x^{(1)}$ y en base a ello determinamos cuál es el que minimiza de mejor manera su valor:

Práctica 2:

Seleccionar una función matemática f(x,y) para realizar el proceso de minimización.

Ejecutar al menos 3 pasos del método de descenso por gradiente.

Diseñar y desarrollar un cuaderno en Jupyter Notebook donde se realicen todos los pasos correspondientes al proceso de minimización de la función a través del método de descenso por gradiente.

El cuaderno deberá incluir los siguientes puntos:

Gráfica de la función y los puntos que se obtienen a medida que se ejecutan los pasos de cálculo (hasta k=3).

Aplicación de las funciones de derivación y evaluación de forma similar a la que se ha detallado en el presente cuaderno.

Incluir un acápite sobre las funciones cóncavas y los puntos estacionarios (incluir gráficos).

Emplear las funcionalidades que proveen los paquetes matplotlib y sympy.

Ejemplo de aplicación

Punto de partida es $x^{(0)}=(3,4)$:

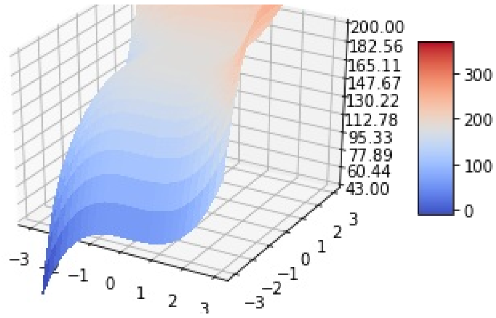
$f(x,y)=3x^{3}+2xy+5y^{3}+8.9$

Importamos librerias

In [39]:

Creamos la malla y la superficie

In [40]:



Descenso por Gradiente

Dado que no conocemos el valor óptimo de x, hacemos $k\leftarrow 0$, como nos indica el algoritmo. A continuación procedemos a calcular los siguientes elementos:

- La derivada **parcial** de la función original $f(x,y)=3x^{3}+2xy+5y^{3}+8.9$.
- Evaluamos la función original en el punto $x^{(0)}=(3,4)$.

Dibujamos la superficie

Importamos librerias

In [41]:

In [42]:

```
[ 9*x**2 + 2*y , 2*x + 15*y**2 ]
Valor en el punto (x^(0),y^(0)) => (3,4) =  433.900000000000
```

Lo que nos da como resultado los siguientes valores:

$\nabla f(x,y) = \binom{\frac{\partial f(x,y)}{\partial x}}{\frac{\partial f(x,y)}{\partial y}}= \binom{9x^{2}+2y}{2x+15y^{2}} $

$\nabla f\left((x^{(0)},y^{(0)}) = (3,4)\right) = 433.9$

1. Ahora buscaremos encontrar el siguiente punto con coordenadas

$\left(x^{(k+1)}, y^{(k+1)}\right)$, para ello debemos calcular:

$\left(x^{(1)},y^{(1)}\right)=(x^{(0)},y^{(0)})-t^{*}_{0} \nabla f(x^{(0)},y^{(0)})$

2. Para encontrar el valor de t^{*}_{0}, debemos hallar el mínimo de la función $\theta(t) = f(x^{(0)},y^{(0)})-t \nabla f(x^{(0),y^{(0)}}$. Para realizar este paso, buscamos en el punto estacionario trabajando con la derivada de la función a minimizar. En este punto, usaremos una notación vectorial:

- Para esto, calculamos el valor de t^{*}_{k} empleando el punto $x^{(0)}$, y la función de t y en base a la derivada de la función a minimizar:

$\theta^{'}(t)=-\nabla f\left((x^{(0)},y^{(0)})- t \nabla f(x^{(0)},y^{(0)}) \right) \nabla f(x^{(0)},y^{(0)})$

$=-\nabla f\left((3,4)- t \nabla f(3,4) \right) \nabla f(x^{(0)},y^{(0)})$

Para hallar $\nabla f(x^{(0)},y^{(0)})$, simplemente sustituimos el punto $(3,4)$ en ambas partes de la función original que se derivó con respecto a x y con respecto a y:

$\nabla f(x^{(0)},y^{(0)}) = \binom{9x^{2}+2y}{2x+15y^{2}}$

$\nabla f(3,4) = \binom{9\cdot 3^{2}+2(4)}{2(3)+15\cdot 4^{2}}$

$\nabla f(3,4) = (89, 246)$

Con ello, ahora volvemos a la función $\theta^{'}(t)$ y reemplazamos el valor calculado:

$=-\nabla f\left((3,4)- t (89,246) \right) (89,246)$

$=-\nabla f\left((3-89t),(4-246t) \right) (89,246)$

Ahora, evaluamos la función que derivamos con los nuevos valores $x=(3-89t), y=(4-246t)$:

$=-\left(9\cdot (3-89t)^{2}+2(4-246t), 2(3-89t)+15\cdot (4-246t)^{2} \right) (89,246)$

$=-\lbrace{ 89\left(9\cdot (3-89t)^{2}+2(4-246t)\right)+246\left(2(3-89t)+15\cdot (4-246t)^{2} \right) \rbrace}$

Factoramos con ayuda de **sympy** y buscamos las raíces:

In [57]:

```
-229648761.0*t**2 + 7777230.0*t - 68437.0
```

Resultados

In [58]:

```
Derivada para buscar raices:  7777230.0 - 459297522.0*t
*****************************************************
Raices: [ [0.0169328803824898] ]
********************************
```

Con ello, nuestra ecuación en función de t queda como sigue:

$=-229648761t^{2} + 7777230t - 68437$

Y la raíz que encontramos luego de derivarla (ya que nos salen valores imaginarios) es $t_{0}=0.0169$.

3. Dado lo anterior, el siguiente punto $(x^{(1)},y^{(1)})$ será:

$(x^{(1)},y^{(1)})=(x^{(0)},y^{(0)})-t^{*}_{0} \nabla f(x^{(0)},y^{(0)})$

$(x^{(1)},y^{(1)})=(3,4)-0.0169\cdot (15, 29)$

$(x^{(1)},y^{(1)})=(2.7465, 3.5099)$

Si ahora evaluamos la función original con el nuevo punto $(x^{(1)},y^{(1)})=(2.7465, 3.5099)$, obtenemos obtendremos lo siguiente:

In [59]:

```
Valor en el punto (x^(1),y^(1)) => (2.7465,3.5099) =  306.531865385370
```

SEGUNDA INTERACCION

Importancion de liberias

In [60]:

Procesamiento

Procesamiento

In [61]:

Resultados

In [63]:

```
[ 9*x**2 + 2*y , 2*x + 15*y**2 ]
Valor en el punto (x^(0),y^(0)) => (2.7465,3.5099) =  306.531865385370
```

$\nabla f(x,y) = \binom{\frac{\partial f(x,y)}{\partial x}}{\frac{\partial f(x,y)}{\partial y}}= \binom{9x^{2}+2y}{2x+15y^{2}} $

$\nabla f\left((x^{(1)},y^{(1)}) = (2.7465, 3.5099)\right) = 306.531865$

$\left(x^{(1)},y^{(1)}\right)=(x^{(1)},y^{(1)})-t^{*}_{1} \nabla f(x^{(1)},y^{(1)})$

- Para esto, calculamos el valor de t^{*}_{k} empleando el punto $x^{(1)}$, y la función de t y en base a la derivada de la función a minimizar:

$\theta^{'}(t)=-\nabla f\left((x^{(1)},y^{(1)})- t \nabla f(x^{(1)},y^{(1)}) \right) \nabla f(x^{(1)},y^{(1)})$

$=-\nabla f\left((2.7465,3.5099)- t \nabla f(2.7465,3.5099) \right) \nabla f(x^{(1)},y^{(1)})$

Para hallar $\nabla f(x^{(1)},y^{(1)})$, simplemente sustituimos el punto $(2.7465,3.5099)$ en ambas partes de la función original que se derivó con respecto a x y con respecto a y:

$\nabla f(x^{(1)},y^{(1)}) = \binom{9x^{2}+2y}{2x+15y^{2}}$

$\nabla f(2.7465,3.5099) = \binom{9\cdot 2.7465^{2}+2(3.5099)}{2(2.7465)+15\cdot 3.5099^{2}}$

$\nabla f(2.7465,3.5099) = (74.91, 190.28)$

Con ello, ahora volvemos a la función $\theta^{'}(t)$ y reemplazamos el valor calculado:

$=-\nabla f\left((2.75,3.51)- t (74.91,190.28) \right) (74.91,190.28)$

$=-\nabla f\left((2.75-74.91t),(3.51-190.28t) \right) (74.91,190.28)$

Ahora, evaluamos la función que derivamos con los nuevos valores $x=(2.75-74.91t), y=(3.51-190.28t)$:

$=-\left(9\cdot (2.75-74.91t)^{2}+2(3.51-190.28t), 2(2.75-74.91t)+15\cdot (3.51-190.28t)^{2} \right) (74.91,190.28$

$=-\lbrace{ 74.91\left(9\cdot (2.75-74.91t)^{2}+2(3.51-190.28t)\right)+190.28\left(2(2.75-74.91t)+15\cdot (3.51-190.28t)^{2} \right) \rbrace}$

Factoramos con ayuda de **sympy** y buscamos las raíces:

In [65]:

```
-107123753.295219*t**2 + 4147327.32567*t - 41834.999495
```

Resultado

In [66]:

```
Derivada para buscar raices:  4147327.32567 - 214247506.590438*t
Raices: [ [0.0193576457045923] ]
```

Con ello, nuestra ecuación en función de t queda como sigue:

$$=-107123753.295219t^{2} + 4147327.32567t - 4147327.33$$

Y la raíz que encontramos luego de derivarla (ya que nos salen valores imaginarios) es $t_{1}=0.0193$.

3. Dado lo anterior, el siguiente punto $(x^{(2)},y^{(2)})$ será:

$$(x^{(2)},y^{(2)})=(x^{(1)},y^{(1)})-t^{*}_{1} \nabla f(x^{(1)},y^{(1)})$$

$$(x^{(2)},y^{(2)})=(2.75,3.51)-0.0193\cdot (74.91, 190.28)$$

$$(x^{(2)},y^{(2)})=(1.3, 3.51)$$

Si ahora evaluamos la función original con el nuevo punto $(x^{(2)},y^{(2)})=(1.3, 3.51)$, obtenemos obtendremos lo siguiente:

In [67]:

```
Valor en el punto (x^(1),y^(1)) => (0.3155,-0.25669) =  240.834755000000
```

Conclusiones

El método del gradiente descendiente es muy usado para entrenar redes neuronales y también en aprendizaje profundo.

En particular, nos interesa utilizar el gradiente descendiente para estimar los parámetros W que minimizan la función de coste. También hemos visto por qué es tan importante escalar los datos antes de utilizar este método de optimización numérica.

Por último, hemos vistos que para un conjunto de datos muy amplio, el gradiente descendiente nos puede interesar más que el uso del método de los mínimos cuadrados.

Referencias

[1] O. Axelsson. Iterative Solution Methods. Ed. Cambridge University Press, 1996.

[2] O. Axelsson, V.A. Barker. Finite Element Solution of Boundary Value Problems. Theory and Computations. Ed. Academic Press, 1984.

[3] R. L. Burden y J. D. Faires. M´etodos Num´ericos (3a Edici´on). Ed. Thomson, 2004.

[4] C. Conde y G. Winter. M´etodos y algoritmos b´asicos del ´algebra num´erica. Ed. Revert´e, 1989

Repositorio

Git:https://github.com/ChristianHernand16/Inteligencia-Artificial-II.git