Matematica 4

TP Analisis matematico y regresión lineal

Integrantes:

Mercuri, LUCIANO NICOLAS

Kessler, ANDRÉS GABRIEL

Introducción

En este TP vamos a trabajar con funciones de varias variables, analizando su dominio, y aspectos como la continuidad, la diferenciabilidad, planos tangentes y sus mínimos y máximos locales/absolutos.
También analizamos el proceso iterativo para encontrar un mínimo local de una función a partir de un punto específico utilizando el método de descenso por gradiente.
Por último estudiamos el método de regresión lineal, principalmente mediante el método de mínimos cuadrados para calcular una recta estimada a partir de datos previamente establecidos, demostrando la influencia del alfa y el coeficiente de determinación, además de graficar a partir de datos reales mediante software (en nuestro caso un programa de Python, utilizando las librerías “matplotlib” y “pandas”).

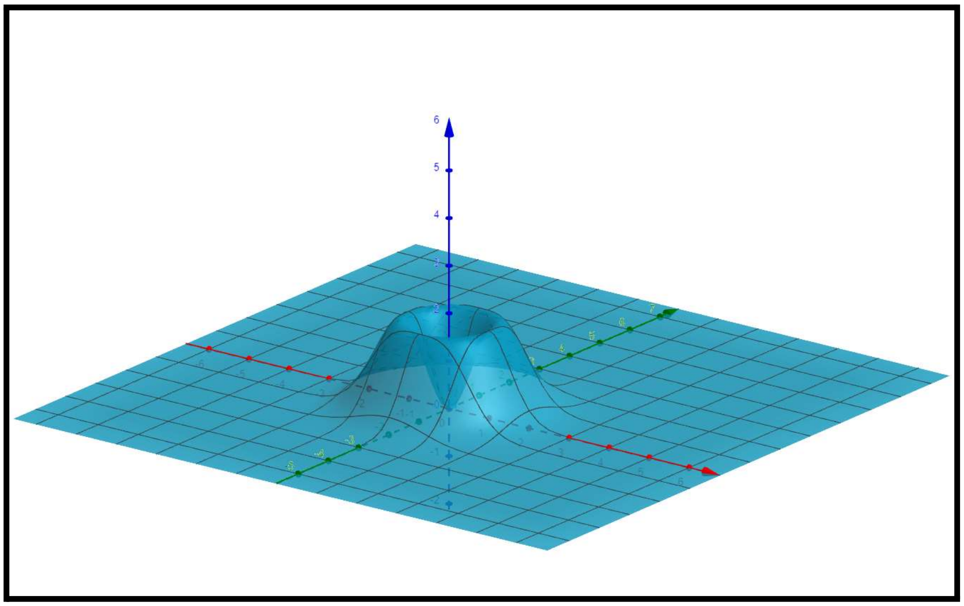
Resolución

Ejercicio 1

Dada la siguiente función:

$$f(x, y) = e^{(-x^2-y^2)}(5x^2 + 5y^2)$$

a) Graficar la función y mostrar el resultado.



b) Analizar su dominio y continuidad en todo R².

Podemos ver que la función *f(x,y)* está definida para todo (x,y) ∊ R^2, por tanto $D = R^2$ y su imagen es *Im(f) = R.*
También podemos ver que la función es continua en todo R^2. Esto lo podemos demostrar usando los siguientes teoremas:

- Una función exponencial es continua en todo R^2. Ej: $e^{(-x^2-y^2)}$
- Una función polinomial es continua en todo R^2. Ej: $(5x^2 + 5y^2)$
- El producto de dos funciones continuas es continua.

c) Estudiar (por definición) la diferenciabilidad de la función en el punto (0, 0). ¿Qué entiende por "función diferenciable"?

Para probar la diferenciabilidad de una función se tiene que cumplir:

$$\frac{f(x,y)-\left\{\frac{\partial f(x_0,y_0)}{\partial x}(x-x_0)+\frac{\partial f(x_0,y_0)}{\partial y}(y-y_0)+f(x_0,y_0)\right\}}{\sqrt{(x-x_0)^2+(y-y_0)^2}} \to 0$$

fig. 1

Primero calculamos las derivadas parcial de x en el punto (0,0).

$$\frac{\partial}{\partial x}\left(e^{\left(-x^2-y^2\right)}\left(5x^2+5y^2\right)\right)= -2e^{-x^2-y^2}x\left(5x^2+5y^2\right)+10e^{-x^2-y^2}x$$

Si reemplazamos (x,y) por (0,0) obtenemos

$$\frac{\partial}{\partial x}\left(e^{\left(-x^2-y^2\right)}\cdot\left(5x^2+5y^2\right)\right)=0$$

Realizamos el mismo proceso para la derivada parcial de y

$$\frac{\partial}{\partial y}\left(e^{\left(-x^2-y^2\right)}\left(5x^2+5y^2\right)\right)= -2e^{-x^2-y^2}y\left(5x^2+5y^2\right)+10e^{-x^2-y^2}y$$

$$\frac{\partial}{\partial y}\left(e^{\left(-x^2-y^2\right)}\cdot\left(5x^2+5y^2\right)\right)=0$$

Ahora podemos reemplazar la fórmula en *fig 1*:

- $$\frac{e^{-x^2-y^2}(5x^2+5y^2)-[0*(x-0)+0*(y-0)+e^{-0^2-0^2}(5*0^2+5*0^2)]}{\sqrt{(x-0)^2+(y-0)^2}}$$

- $$\frac{e^{-x^2-y^2}(5x^2+5y^2)-0}{\sqrt{x^2+y^2}}$$

- $\frac{e^{-x^2-y^2}5(x^2+y^2)}{\sqrt{x^2+y^2}}$

- $e^{-x^2-y^2}5\sqrt{x^2+y^2}$

Podemos ver que si (x,y) → (0,0) entonces

$$e^{-x^2-y^2}5\sqrt{x^2+y^2} \to 0$$

Por lo tanto comprobamos que f(x,y) es diferenciable en punto (0,0).
Se entiende por diferenciabilidad cuando la función es indistinguible del plano tangente en el punto (x0,y0).

d) ¿Es diferenciable en todo R^2?

Si, es diferenciable en todo R^2. Esto lo podemos probar mediante el teorema:

> *Sea $f : D \subset \mathbb{R}^2 \to \mathbb{R}$ y sea $(x_0, y_0) \in D$. Si existen las derivadas parciales $\frac{\partial f(x,y)}{\partial x}$ y $\frac{\partial f(x,y)}{\partial y}$, y además éstas son continuas en un entorno del punto (x_0, y_0), entonces f es diferenciable en (x_0, y_0).*

Sabemos que las derivadas parciales son continuas

$$\frac{\partial f(x,y)}{\partial x} = -2e^{-x^2-y^2} * x * (5x^2 + 5y^2) + 10e^{-x^2-y^2} * x$$

$$\frac{\partial f(x,y)}{\partial y} = -2e^{-x^2-y^2} * y * (5x^2 + 5y^2) + 10e^{-x^2-y^2} * y$$

Como ambas son continuas en todo punto de R^2, f es diferenciable en todo R^2.

e) ¿La función tiene algún máximo absoluto? Si existe, ¿Éste es único? ¿A qué se debe?

Si existe, pero no es único. Esto lo podemos demostrar de la siguiente manera.
Primero hallamos los puntos críticos igualando las derivadas parciales a 0.

$$-2e^{-x^2-y^2} * x * (5x^2 + 5y^2) + 10e^{-x^2-y^2} * x = 0$$

$$-2e^{-x^2-y^2} * x * (5x^2 + 5y^2) + 10e^{-x^2-y^2} * x = 0$$

Despejando obtenemos estos resultados:

$$x = 0 \qquad x = \pm\sqrt{-y^2 + 1}$$

$$y = 0 \qquad y = \pm\sqrt{-x^2 + 1}$$

De esta forma encontramos los puntos críticos en (0,0) y en las infinitas iteraciones de $(x, \sqrt{-x^2+1})$ y $(\sqrt{-y^2+1}, y)$.

Sabemos que f$(x, \sqrt{-x^2+1})$ ≥ f(x,y), al igual que f$(\sqrt{-y^2+1}, y)$ ≥ f(x,y), por ende en todos esos puntos hay un máximo absoluto. Notar que 'x' o 'y' tienen que pertenecer al intervalo [-1;1] y se podría decir que hay una "circunferencia de máximos absolutos" con radio igual a 1.

f) ¿El punto (0, 0) analizado, es un punto crítico? Si lo es, ¿de qué tipo?

En el punto anterior encontramos que el punto (0,0) es un punto crítico ya que ambas derivadas parciales se anulan cuando x = 0, y = 0.
Como f(0,0) ≤ f(x,y) en todo el dominio de f, se puede deducir que tiene un mínimo absoluto en (0,0)

g) Calcule, si es posible, el plano tangente a la función en el punto (0.5, 1). En el caso de que sea posible, añádelo al gráfico del inciso (a).

Tenemos la siguiente ecuación del plano tangente:

$$z = \frac{\partial f(x0,y0)}{\partial x}(x - x0) + \frac{\partial f(x0,y0)}{\partial y}(y - y0) + f(x0, y0)$$

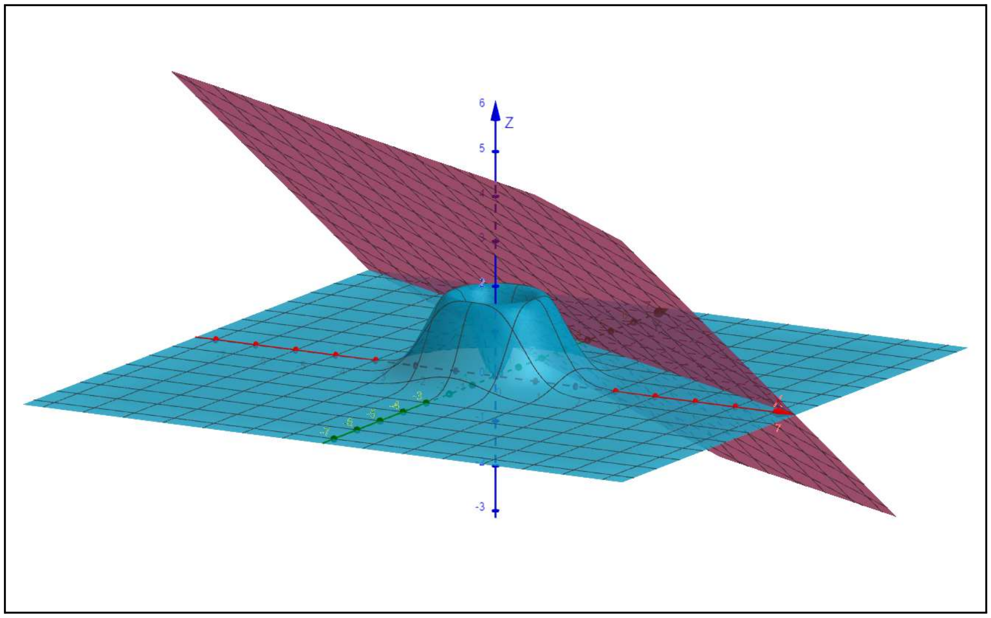
También tenemos los siguientes datos, calculados en el punto (0.5 , 1):

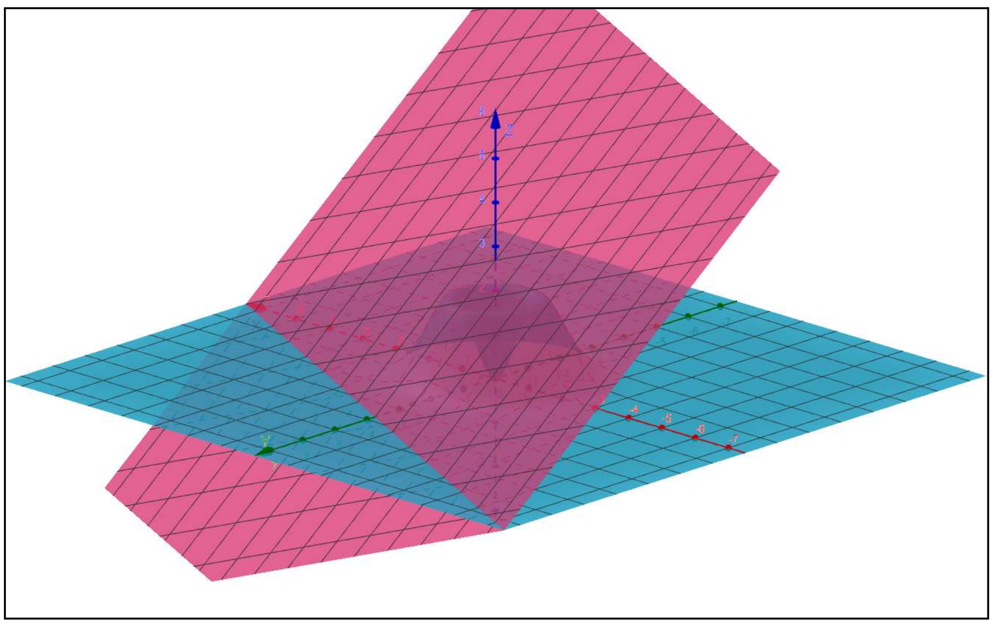
- $\frac{\partial f(x0,y0)}{\partial x} = -2e^{-0.5^2-1^2} * 0.5 * (5 * 0.5^2 + 5 * 1^2) + 10e^{-0.5^2-1^2} * 0.5 = -0.35813$
- $\frac{\partial f(x0,y0)}{\partial y} = -2e^{-0.5^2-1^2} * 1 * (5 * 0.5^2 + 5 * 1^2) + 10e^{-0.5^2-1^2} * 1 = -0.71626$
- $f(0.5, 1) = e^{-0.5^2-1^2}(5 * 0.5^2 + 5 * 1^2) = 1.79065$

Ahora reemplazamos en la ecuación original

$$z = -0.35813(x - 0.5) - 0.71626(y - 1) + 1.79065$$

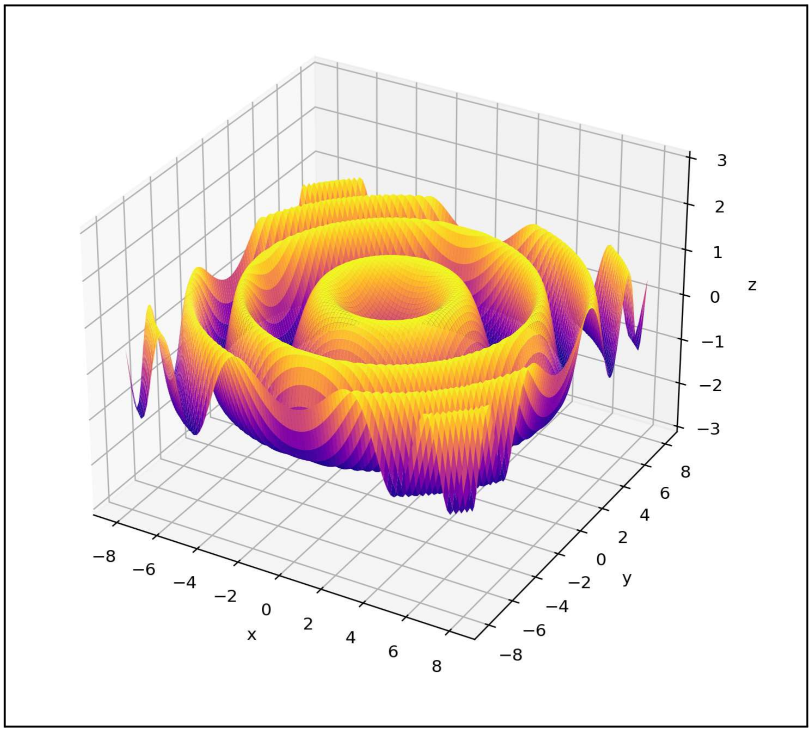
$$z = 2.685975 - 0.35813x - 0.71626y$$





Ejercicio 2

a) Graficar la función (tomar el intervalo [-8,8] tanto para x como para y) y mostrar el resultado.



b) Desarrollar (por escrito) 3 iteraciones del método implementado, partiendo del punto (3, 0, g(3, 0)) y utilizando α = 0.5. Dar los valores de las variables utilizadas en cada iteración (debe estar presente los valores de las derivadas parciales). ¿El valor final obtenido parece ser un mínimo absoluto de la función? Compare el resultado con el gráfico del inciso anterior.

Datos iniciales:

- $\alpha = 0.5$
- $f(x, y) = sen(\frac{x^2+y^2}{5})$
- $\frac{\partial f(x,y)}{\partial x} = \frac{2*x*cos*(\frac{x^2+y^2}{5})}{5}$
- $\frac{\partial f(x,y)}{\partial y} = \frac{2*y*cos*(\frac{x^2+y^2}{5})}{5}$
- Trabajamos en el punto (3, 0)

Primero determinamos el punto inicial

$$z = sen(\frac{3^2+0^2}{5})$$

$$z = sen(\frac{9}{5})$$

$$z = 0.03141$$

Ahora realizamos las iteraciones, restando a x0 e y0 sus derivadas parciales respectivas multiplicadas por el factor α.

1º Iteración:

$$x1 = x - (\alpha * \frac{2*x*cos*(\frac{x^2+y^2}{5})}{5})$$

$$x1 = 3 - (0.5 * \frac{2*3*cos*(\frac{3^2+0^2}{5})}{5})$$

$$x1 = 3 - (0.5 * 1.1994)$$

$$x1 = 2.4$$

$$y1 = y - (\alpha * \frac{2*y*cos*(\frac{x^2+y^2}{5})}{5})$$

$$y1 = 0 - (0.5 * \frac{2*0*cos*(\frac{3^2+0^2}{5})}{5})$$

$$y1 = 0 - (0.5 * 0)$$

$$y1 = 0$$

$$z = sen(\frac{2.4^2+0^2}{5})$$

$$z = sen(\frac{5.76}{5})$$

$$z = 0.0201$$

2º Iteración:

$$x2 = x1 - (\alpha * \frac{2*x1*cos*(\frac{x1^2+y1^2}{5})}{5})$$

$$x2 = 2.4 - (0.5 * \frac{2*2.4*cos*(\frac{2.4^2+0^2}{5})}{5})$$

$$y2 = y1 - (\alpha * \frac{2*y*cos*(\frac{x1^2+y1^2}{5})}{5})$$

$$y2 = 0 - (0.5 * \frac{2*0*cos*(\frac{2.4^2+0^2}{5})}{5})$$

$x2 = 2.4 - (0.5 * 0.9598)$ $\quad$ $y1 = 0 - (0.5 * 0)$

$x2 = 1.920$ $\quad$ $y1 = 0$

$$z = sen(\frac{1.920^2 + 0^2}{5})$$

$$z = sen(\frac{3.6867}{5})$$

$$z = 0.0128$$

3º Iteración:

$x3 = x2 - (\alpha * \frac{2*x2*cos*(\frac{x2^2+y2^2}{5})}{5})$ $\quad$ $y2 = y2 - (\alpha * \frac{2*y2*cos*(\frac{x2^2+y2^2}{5})}{5})$

$x3 = 1.92 - (0.5 * \frac{2*1.92*cos*(\frac{1.92^2+0^2}{5})}{5})$ $\quad$ $y3 = 0 - (0.5 * \frac{2*0*cos*(\frac{1.92^2+0^2}{5})}{5})$

$x3 = 1.92 - (0.5 * 0.7679)$ $\quad$ $y3 = 0 - (0.5 * 0)$

$x3 = 1.536$ $\quad$ $y3 = 0$

$$z = sen(\frac{1.536^2 + 0^2}{5})$$

$$z = sen(\frac{2.359}{5})$$

$$z = 0.0082$$

A primera vista parece que se está llegando a un mínimo en $z = 0$, pero al ver la gráfica de la función, cuando iniciamos en el punto (3,0), la función converge en ~(-1). Esta discrepancia puede ser debido a que 3 iteraciones no son suficiente para acercarnos al punto de convergencia o nuestro α no es apropiado.

c) Realice los mismos pasos que en el punto anterior pero incrementando el valor de α a 0.8. ¿Se alcanzó el mínimo o estuvo cerca de hacerlo?

1º Iteración:

$x1 = x - (\alpha * \frac{2*x*cos*(\frac{x^2+y^2}{5})}{5})$ $y1 = y - (\alpha * \frac{2*y*cos*(\frac{x^2+y^2}{5})}{5})$

$x1 = 3 - (0.8 * \frac{2*3*cos*(\frac{3^2+0^2}{5})}{5})$ $y1 = 0 - (0.8 * \frac{2*0*cos*(\frac{3^2+0^2}{5})}{5})$

$x1 = 3 - (0.8 * 1.1994)$ $y1 = 0 - (0.8 * 0)$

$x1 = 2.04$ $y1 = 0$

$$z = sen(\frac{2.04^2+0^2}{5})$$

$$z = sen(\frac{2.04^2}{5})$$

$$z = 0.0145$$

2º Iteración:

$x2 = x1 - (\alpha * \frac{2*x1*cos*(\frac{x1^2+y1^2}{5})}{5})$ $y2 = y1 - (\alpha * \frac{2*y*cos*(\frac{x1^2+y1^2}{5})}{5})$

$x2 = 2.04 - (0.8 * \frac{2*2.04*cos*(\frac{2.04^2+0^2}{5})}{5})$ $y2 = 0 - (0.8 * \frac{2*0*cos*(\frac{2.04^2+0^2}{5})}{5})$

$x2 = 2.04 - (0.8 * 0.8159)$ $y1 = 0 - (0.8 * 0)$

$x2 = 1.387$ $y1 = 0$

$$z = sen(\frac{1.387^2+0^2}{5})$$

$$z = sen(\frac{1.387^2}{5})$$

$$z = 0.0067$$

3º Iteración:

$$x3 = x2 - (\alpha * \frac{2*x2*cos*(\frac{x2^2+y2^2}{5})}{5}) \qquad y2 = y2 - (\alpha * \frac{2*y2*cos*(\frac{x2^2+y2^2}{5})}{5})$$

$$x3 = 1.387 - (0.8 * \frac{2*1.387*cos*(\frac{1.387^2+0^2}{5})}{5}) \qquad y3 = 0 - (0.8 * \frac{2*0*cos*(\frac{1.387^2+0^2}{5})}{5})$$

$$x3 = 1.387 - (0.8 * 0.5547) \qquad y3 = 0 - (0.8 * 0)$$

$$x3 = 0.9431 \qquad y3 = 0$$

$$z = sen(\frac{0.9431^2+0^2}{5})$$

$$z = sen(\frac{0.9431^2}{5})$$

$$z = 0.0031$$

De forma similar al caso anterior, parece que converge en $z = 0$, pero también es erróneo. Una diferencia respecto al caso anterior sería que si aumentamos bastante la cantidad de iteraciones posibles (y tomamos una tolerancia satisfactoria) en ambos casos, en el primer caso la función converge en -1. En cambio en el segundo caso la función se queda “rebotando” sin alcanzar la tolerancia, esto se debe a que el α es muy grande.

d) ¿Qué función cumple el coeficiente α en el método? ¿Podemos asignarle cualquier valor? ¿Si se eligiera un valor muy grande, cómo afectará al desarrollo del método?

El coeficiente α cumple la función de desplazarnos en la dirección del gradiente negativo.
No podemos asignarle cualquier valor, ya que si es muy grande no se estabiliza en un punto específico (se queda “rebotando”), y si es muy chico tardamos demasiado en llegar al mínimo, gastando todas las iteraciones disponibles.

e) ¿Qué función cumple la tolerancia en el cálculo?

La tolerancia cumple la función de saber si hay que seguir iterando. Si la diferencia entre los resultados de dos iteraciones consecutivas es igual o menor que la tolerancia fijada podemos decir que se “estabiliza” y no hace falta reiterar más.

f) Ejecutar dos veces el método implementado con los siguientes parámetros iniciales:

- **Caso 1:**
  - • **Punto inicial: (3,0,g(3,0))**
  - • **Tolerancia: 10^−5**
  - • **Máximas iteraciones: 50**
  - • **Valor de α: 0.2**

- **Caso 2:**
  - • **Punto inicial: (1,1,g(1,1))**
  - • **Tolerancia: 10^−5**
  - • **Máximas iteraciones: 50**
  - • **Valor de α: 0.2**

Responder:
i. ¿Se utilizó el máximo de iteraciones para alcanzar el valor aproximado?

En el primer caso solo hizo falta 14 iteraciones, en cambio en el caso 2 se llegó al máximo de iteraciones.

ii. ¿Se alcanzó el mínimo absoluto en ambos casos? ¿A qué se debe?

No. Esto se debe a que estamos iniciando el método en distintos puntos. En el caso 1 converge en el -1, mientras que en el caso 2 converge en el 0.

iii. ¿El método siempre converge al mínimo absoluto de la función?

No, distintos puntos pueden tener distintas direcciones de máximo crecimiento (o decrecimiento), lo cual nos indica un máximo o mínimo local, no es necesariamente absoluto.

Ejercicio 3

Un virus informático atacó el disco duro (HDD) de una computadora y a partir de una investigación realizada por expertos sobre el tema se adoptó el siguiente modelo Y = 10β1x + β0 + ϵ para determinar el porcentaje da˜no producido por el virus en función de los d´ıas y se obtuvieron los siguientes datos:

$$S_{xx} = 27768.357 \quad \bar{x} = 97.2143 \quad \hat{\beta}_0 = -0.4298$$

$$\sum {x_i}^2 = 160077 \qquad \sum x_i y_i = 9885 \qquad \sum {y_i}^2 = 27768.357$$

a) Calcular los estimadores de β1 y β0 por el método de mínimos cuadrados, detallando cada uno de los pasos.

Primero hay que intentar despejar la mayor cantidad de datos que podamos utilizando las siguientes fórmulas:

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\sum x_i \sum y_i}{n} \qquad S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$\hat{\beta}_1 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_0 = \frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n} = \overline{y} - \hat{\beta}_1 \overline{x}$$

Lo primero que podemos despejar es $\sum xi$:

$$Sxx = \sum x^2 - \frac{(\sum xi)^2}{n}$$

$$27768.357 = 160077 - \frac{(\sum xi)*(\sum xi)}{n}$$

$$27768.357 = 160077 - \frac{(\sum xi)}{n} * \sum xi$$

$$27768.357 = 160077 - \overline{x} * \sum xi$$

$$27768.357 = 160077 - 97.2143 * \sum xi$$

$$\frac{27768.357 - 160077}{-97.2143} = \sum xi$$

$$1361 = \sum xi$$

Luego fácilmente podemos despejar ‘n’:

$$\bar{x} = \frac{\sum xi}{n}$$

$$n = \frac{\sum xi}{\bar{x}}$$

$$n = \frac{1361}{97.2143}$$

$$n = 14$$

Ahora podemos despejar Sxy y $\sum y$ construyendo un sistema de 2 ecuaciones:

$$Sxy = \sum x * y - \frac{\sum x * \sum y}{n}$$

$$\widehat{\beta 0} = \frac{\sum y - \widehat{\beta 1} * \sum x}{n}$$

$$Sxy = 9885 - \frac{1361}{14} * \sum y$$

$$-0.4298 = \frac{\sum y - \frac{Sxy}{27768.357} * 1361}{14}$$

Primero despejamos $\sum y$

$$-0.4298 = \frac{\sum y - \frac{9885 - \frac{1361}{14} * \sum y}{27768.357} * 1361}{14}$$

$$-6.0172 = \sum y - \left(\frac{9885 - 97.2143 * \sum y}{27768.357}\right) * 1361$$

$$-6.0172 = \sum y - \left(\frac{9885}{27768.357} - \frac{97.2143 * \sum y}{27768.357}\right) * 1361$$

$$-6.0172 = \sum y - (0.35598 - 0.0035 * \sum y) * 1361$$

$$-6.0172 = \sum y - 484.4887 + 4.7635 * \sum y$$

$$-6.0172 = -484.4887 + 5.7635 * \sum y$$

$$\frac{-6.0172+484.4887}{5.7635} = \sum y$$

$$83 = \sum y$$

Ahora despejamos Sxy

$$Sxy = 9885 - \frac{1361}{14} * \sum y$$

$$Sxy = 9885 - \frac{1361}{14} * 83$$

$$Sxy = 1814.509$$

Por último, despejamos $\widehat{\beta 1}$:

$$\widehat{\beta 1} = \frac{Sxy}{Sxx}$$

$$\widehat{\beta 1} = \frac{1814.509}{27768.357}$$

$$\widehat{\beta 1} = 0.06534$$

b) Estime la recta de regresión estimada.

$$\hat{Y} = 10 * \widehat{\beta 1} * x + \widehat{\beta 0}$$
$$\hat{Y} = 10 * 0.06534 * x - 0.4298$$
$$\hat{Y} = 0.6534 * x - 0.4298$$

c) ¿En cuántos días se pronostica un daño mayor al 90%?

$$\widehat{Y} = 10 * \widehat{\beta 1} * x + \widehat{\beta 0}$$

$$0.9 = 10 * 0.06534 * x - 0.4298$$

$$0.9 = 0.6534 * x - 0.4298$$

$$\frac{0.9+0.4298}{0.6534} = x$$

$$2.035 = x$$

Se pronostica un daño mayor al 90% en poco más de 2 días.

d) ¿Es posible que la recta verdadera pase por el origen? Realice una prueba de hipótesis adecuada con nivel de significancia 0.05 y responda.

Tomando al origen como (0,0) realizamos el siguiente test de hipótesis:

$H0 : \beta 0 = 0$ $\qquad\qquad$ $H1 : \beta 0 \neq 0$

$$T = \frac{\widehat{\beta 0} - 0}{\sqrt{\widehat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{Sxx}\right)}}$$

Nos hace falta el valor de $\widehat{\sigma}^2$:

$$\widehat{\sigma}^2 = \frac{Syy - \frac{Sxy^2}{Sxx}}{n-2}$$

$$\widehat{\sigma}^2 = \frac{\left(\sum y^2 - \frac{(\sum y)^2}{n}\right) - \frac{Sxy^2}{Sxx}}{n-2}$$

$$\widehat{\sigma}^2 = \frac{\left(27768.357 - \frac{(83)^2}{14}\right) - \frac{1814.509^2}{27768.357}}{14-2}$$

$$\widehat{\sigma}^2 = 2263.143$$

Ahora podemos encontrar nuestro estadístico de prueba:

$$t0 \;=\; \frac{-0.4298\;-\;0}{\sqrt{2263.143\left(\frac{1}{14}+\frac{97.2143}{27768.357}^{2}\right)}}$$

$$t0 \;=\; -\;0.01407$$

Recordando la regla de decisión de la prueba tenemos:

$$Regla: \begin{cases} \text{rechazar} \;\; H_0 \quad si \;\; |T| > t_{\frac{\alpha}{2},\, n-2} \\ \text{aceptar} \;\; H_0 \quad si \;\; |T| \le t_{\frac{\alpha}{2},\, n-2} \end{cases}$$

Buscando en la tabla de distribución Student (con α = 0.05) obtenemos:

$$t\,\frac{0.05}{2},\;14\;-\;2\;=\;2.179$$

Como $-\;0.01407\;\le\;2.179$ no podemos rechazar H0, por lo tanto se podría decir que es posible que la recta pase por el origen.

e) Si del intervalo de confianza para la respuesta media se conoce L2 = 100 (límite superior) y cometiendo un error de 0.05 en el mismo ¿Sobre que dia fijo se realizó el intervalo?

Sabemos el el intervalo de confianza para la respuesta media es:

$$\left[\hat{\beta}_0+\hat{\beta}_1x^*-t_{\frac{\alpha}{2},n-2}\sqrt{\sigma^2\left(\frac{1}{n}+\frac{(x^*-\overline{x})^2}{S_{xx}}\right)}\;;\;\hat{\beta}_0+\hat{\beta}_1x^*+t_{\frac{\alpha}{2},n-2}\sqrt{\sigma^2\left(\frac{1}{n}+\frac{(x^*-\overline{x})^2}{S_{xx}}\right)}\right]$$

Y que el límite superior es igual a 100 y α es igual a 0.05

$$[?,\;100]$$

Por lo tanto

$$100 = \widehat{\beta 0} + \widehat{\beta 1} * x + t_{0.025\,,\,12} * \sqrt{\widehat{\sigma^2} * (1/14 + \frac{(x - \bar{x})^2}{Sxx})}$$

$$100 = -0.4298 + 0.06534 * x + 2.179 * \sqrt{2263.143 * (1/14 + \frac{(x - 97.2143)^2}{27768.357})}$$

Despejando x nos queda

$$x = 30.9462$$

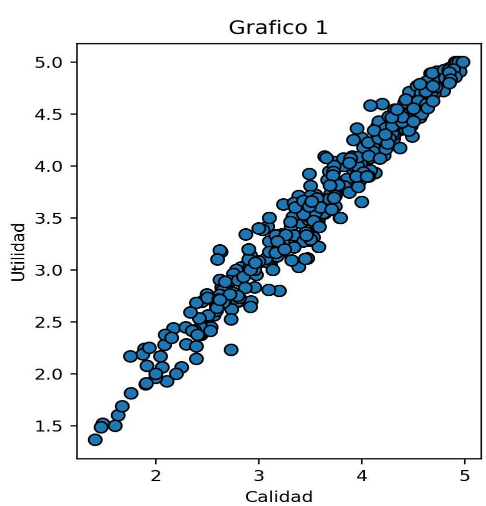
Se podría decir que el intervalo se realizó cerca del día 31.

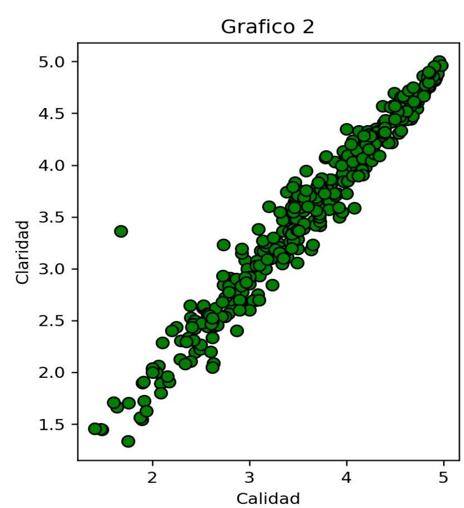
Ejercicio 4

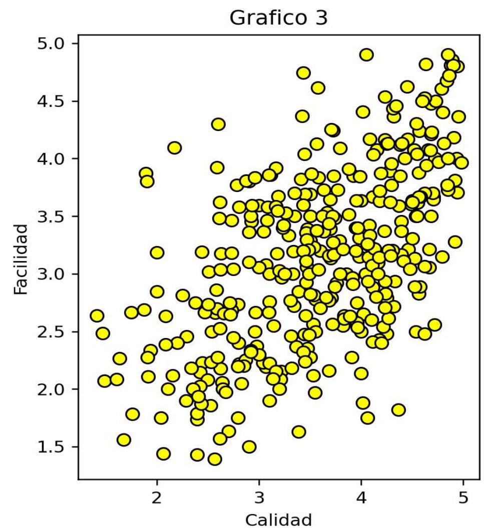
Tomando el dataset ratings.csv realizar 3 análisis de regresión lineal. En todos los casos tomar al atributo calidad como variable independiente o pronosticadora y al resto de atributos como variables dependientes.

Para cada uno de los casos:

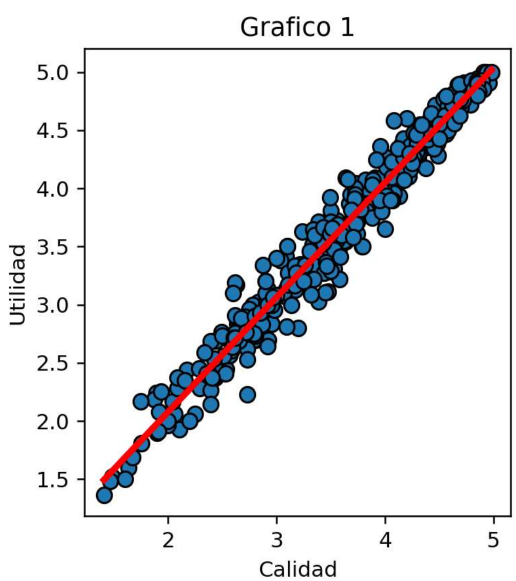
- **Construir y mostrar un gráfico de dispersión.**

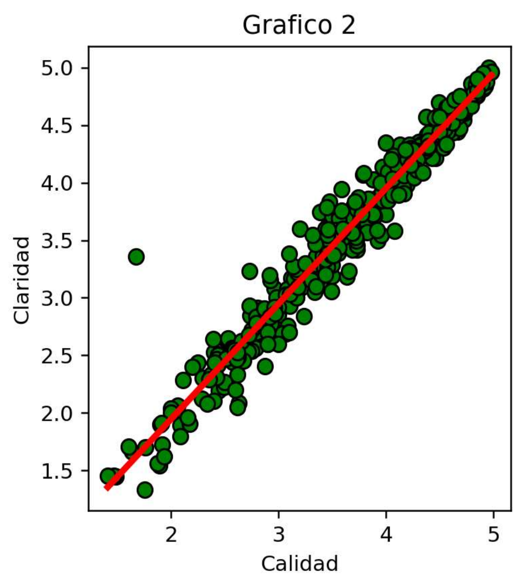


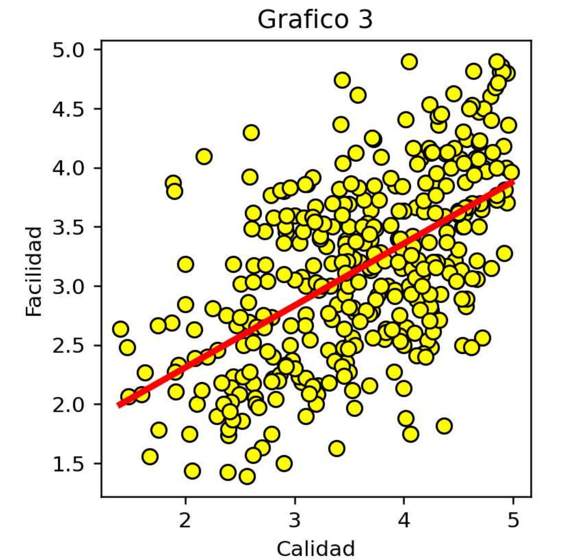




- **Estimar la recta de regresión de mínimos cuadrados y graficarla junto a los puntos del inciso anterior**







- **Calcular la varianza y coeficiente de determinación**

$\sigma^2_1 = 0.02673$ $R_1 = 0.96242 \approx 96\%$

$\sigma^2_2 = 0.03513$ $R_2 = 0.95249 \approx 95\%$

$\sigma^2_3 = 0.41281$ $R_2 = 0.31935 \approx 31\%$

- **Finalmente, observando lo obtenido en cada caso, ¿existe alguna diferencia entre los valores de cada escenario?, ¿ A qué se debe?**

Se nota una diferencia entre cada escenario, principalmente como la varianza influye en el coeficiente de determinación, osea que tan buena es nuestra recta estimada. Por ejemplo, en el 3º caso, debido a la varianza, sólo un 31% de los puntajes en la calidad de las clases es explicada por la facilidad de las mismas.