

Tarea 2

Integrantes:

Salomón Ramírez Quirós - 2017239293

Carlos Barrantes Castro - 2017167999

Curso:

Aprendizaje Automático

I semestre, 2022

4. Derivaciones teóricas:

① Se desea encontrar $\nabla_{\theta} J(\theta)$ e igualarlo a 0 para encontrar una versión generalizada de las ecuaciones normales en este contexto de ponderación:

$$\rightarrow J(\theta) = \frac{1}{2} (X\theta - y)^T W (X\theta - y) \quad ; \quad (AB)^T = B^T A^T$$

$$= \frac{1}{2} \left[\overset{①}{\theta^T X^T W X \theta} + \overset{②}{\theta^T X^T W y} - \overset{③}{y^T W X \theta} + \overset{④}{y^T W y} \right]$$

* Con: $\theta^{m \times 1}$, $X^{n \times m}$, $W^{n \times n}$ y $y^{n \times 1}$

Utilizando propiedades de trazas:

$$① \nabla_{\theta} \theta^T X^T W X \theta = X^T W X \theta + X^T W^T X \theta$$

$$\frac{\partial \text{Tr}(X^T B X)}{\partial X} = B X + B^T X$$

$$② \nabla_{\theta} -\theta^T X^T W y = -X^T W y$$

$$-\frac{\partial \text{Tr}(X^T A)}{\partial X} = -A$$

$$③ \nabla_{\theta} -y^T W X \theta = -X^T W^T y$$

$$-\frac{\partial \text{Tr}(A^T X)}{\partial X} = -A$$

$$④ \nabla_{\theta} y^T W y = 0$$

$$\nabla_{\theta} J(\theta) = [X^T W X \theta + X^T W^T X \theta - X^T W y - X^T W^T y] \cdot \frac{1}{2}$$

→ W es diagonal $\Rightarrow W = W^T$

$$\therefore \nabla_{\theta} J(\theta) = \frac{1}{2} [2X^T W X \theta - 2X^T W y] = X^T W X \theta - X^T W y$$

→ Para minimizar la función de error se iguala $\nabla_{\theta} J(\theta) = 0$

$$X^T W X \theta - X^T W y = 0$$

$$X^T W X \theta = X^T W y$$

$$\theta = (X^T W X)^{-1} X^T W y$$

2-) Verosimilitud: suponga que se tiene.

$$\rightarrow p(y^{(i)} | x^{(i)}; \theta) = \frac{1}{\sqrt{2\pi}\sigma^{(i)}} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right)$$

Demuestre que encontrar el estimado de máxima verosimilitud de θ se reduce a resolver un problema de regresión lineal ponderada e indique claramente cómo se relacionan los pesos $w^{(i)}$ en términos de $\sigma^{(i)}$

→ la verosimilitud se define como:

$$L(\theta) = \prod_{i=1}^m p(y^{(i)} | x^{(i)}; \theta) = \prod_{i=1}^m \frac{1}{\sqrt{2\pi}\sigma^{(i)}} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right)$$

↳ Ya que el producto de muchas probabilidades es numéricamente inestable ($\rightarrow 0$), se utiliza la propiedad \ln

$$J(\theta) = \ln L(\theta) = \sum_{i=1}^m \ln\left(\frac{1}{\sqrt{2\pi}\sigma^i}\right) - \sum_{i=1}^m \frac{(y^i - \theta^T x^i)^2}{2(\sigma^i)^2}$$

Encontrar la máxima verosimilitud $L(\theta)$ es lo mismo que minimizar $J(\theta)$. Si relacionamos la segunda sumatoria de

$J(\theta)$ con $J(\theta)$ tenemos:

$$w^i = \frac{-1}{2(\sigma^i)^2}$$

Reporte de las gráficas obtenidas en la tarea 2.

La figura 1 muestra el 0.1% de los datos censados, los cuales se utilizan para generar los parámetros theta para cada uno de los métodos.

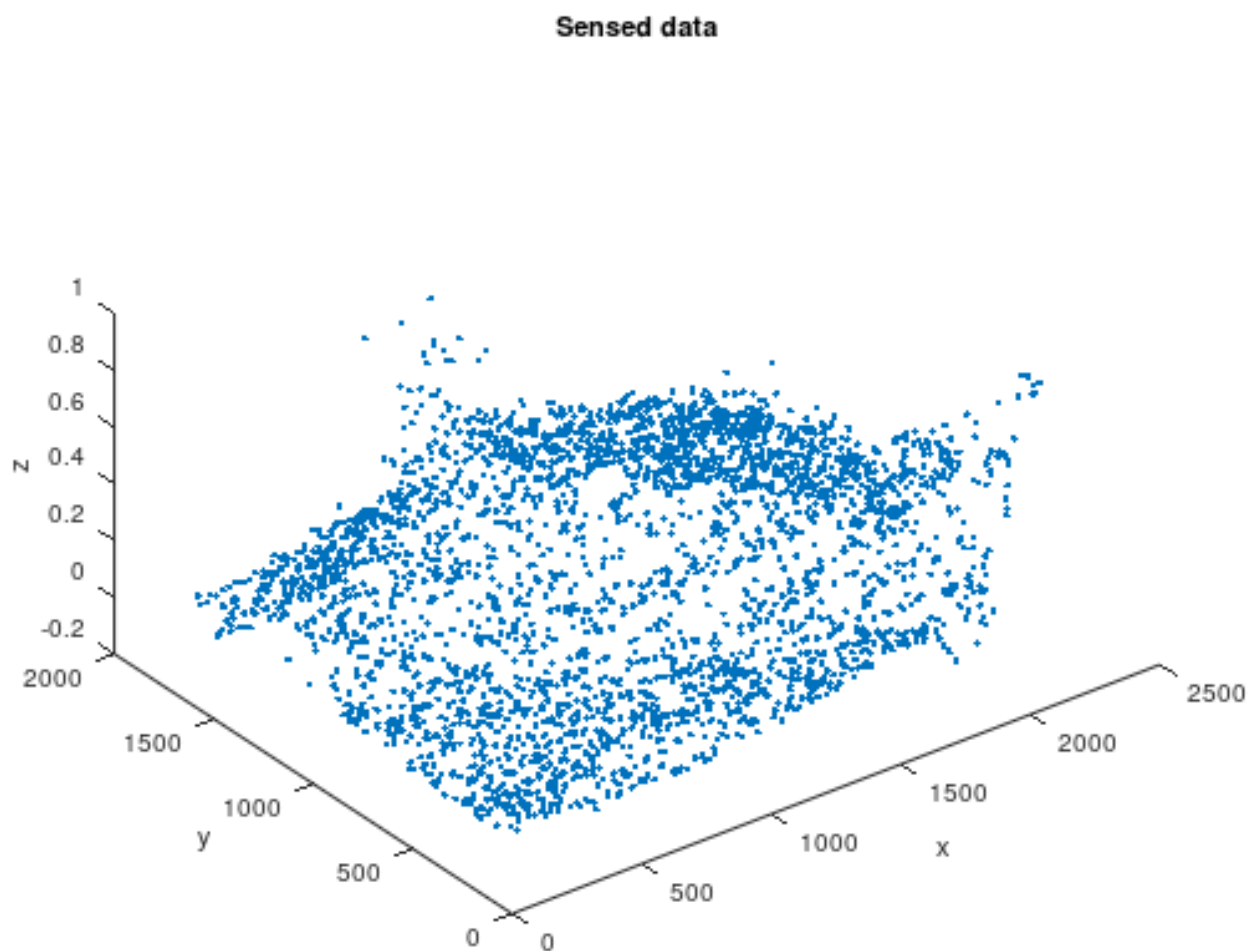


Figura1: 0.1% de los datos censados.

La figura 2 muestra un plano que es la aproximación lineal de la superficie, pero en este caso no se extendió la matriz de diseño con la columna de "1" por lo que no hay sesgo en este plano, lo que carece de sentido físico pero se utiliza como punto de partida.

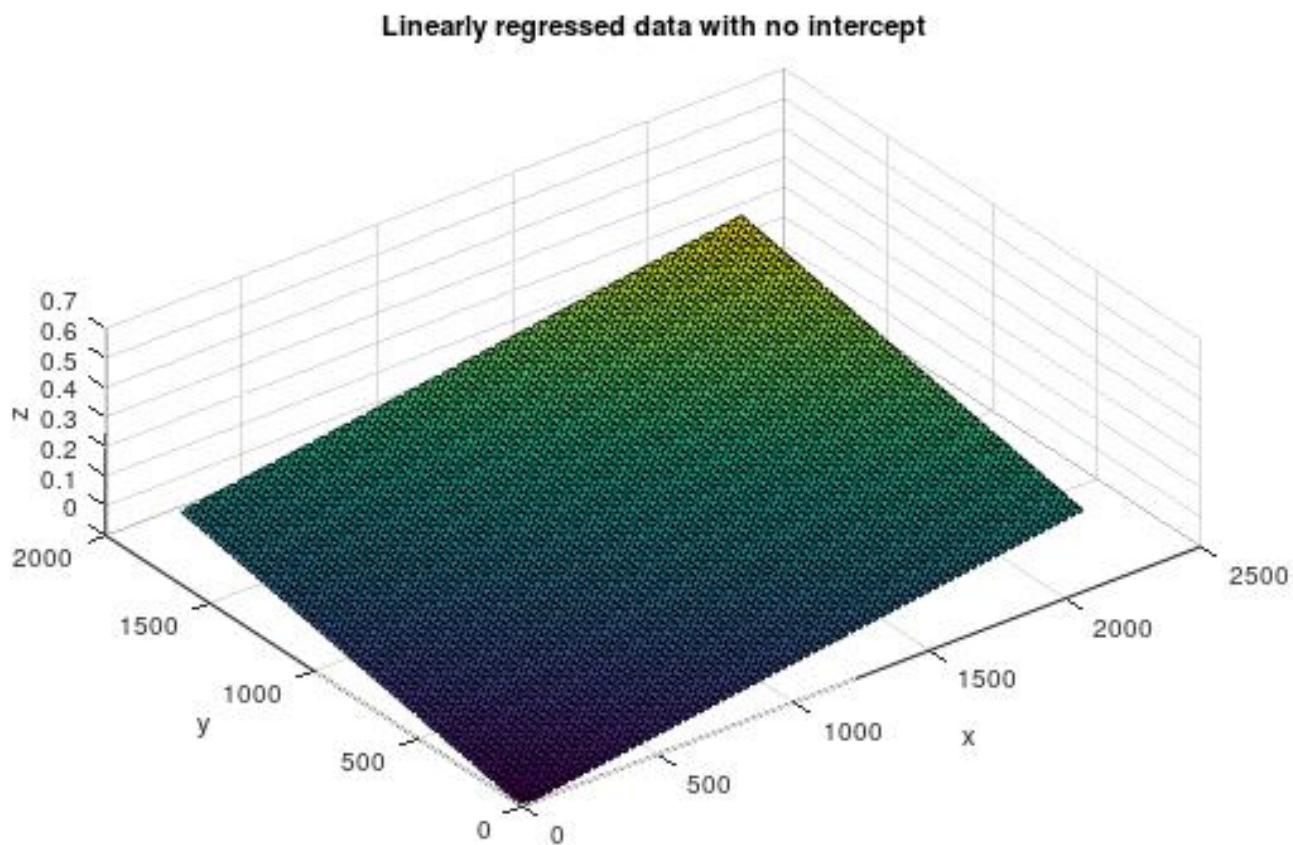


Figura 2: Regresión lineal sin sesgo.

La figura 3 muestra la aproximación lineal de la superficie, esta vez con sesgo.

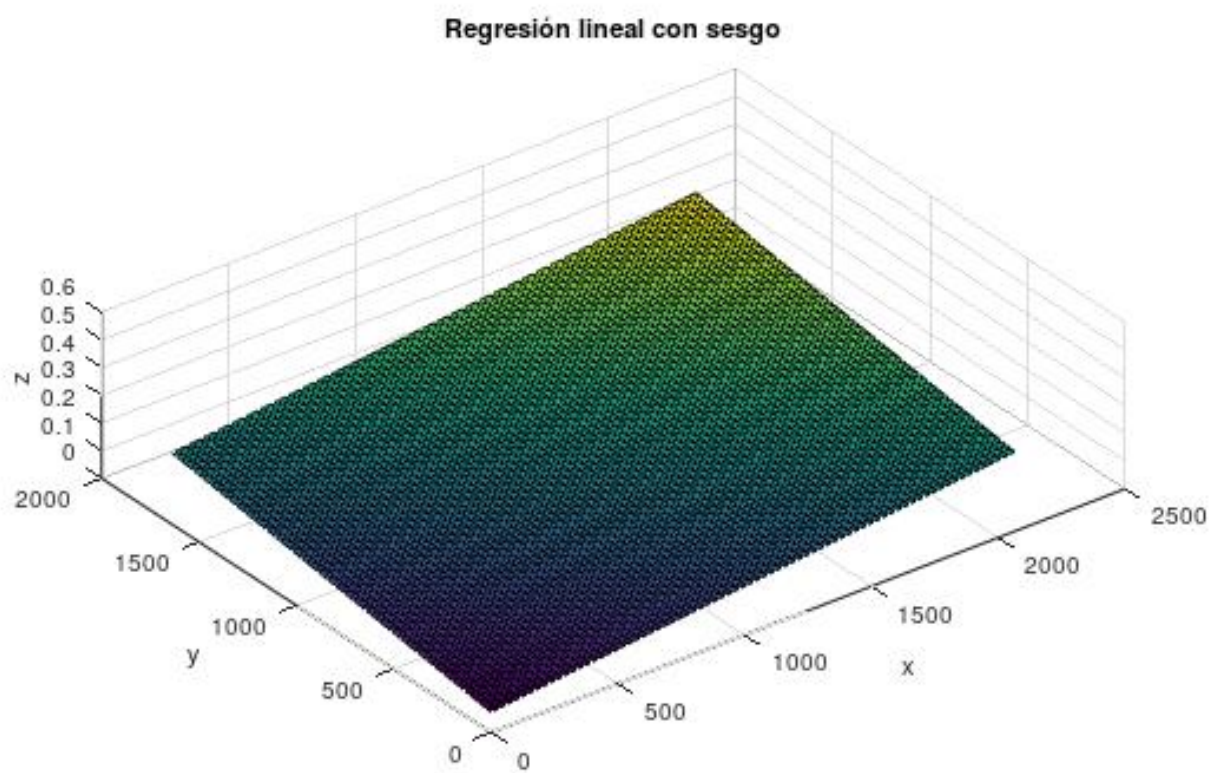


Figura 3: Regresión lineal con sesgo.

La figura 4 muestra una aproximación a la superficie utilizando el método de regresión polinomial igualmente con sesgo, en el script "regressall" se puede definir el grado del polinomio que se desea utilizar para generar la aproximación.

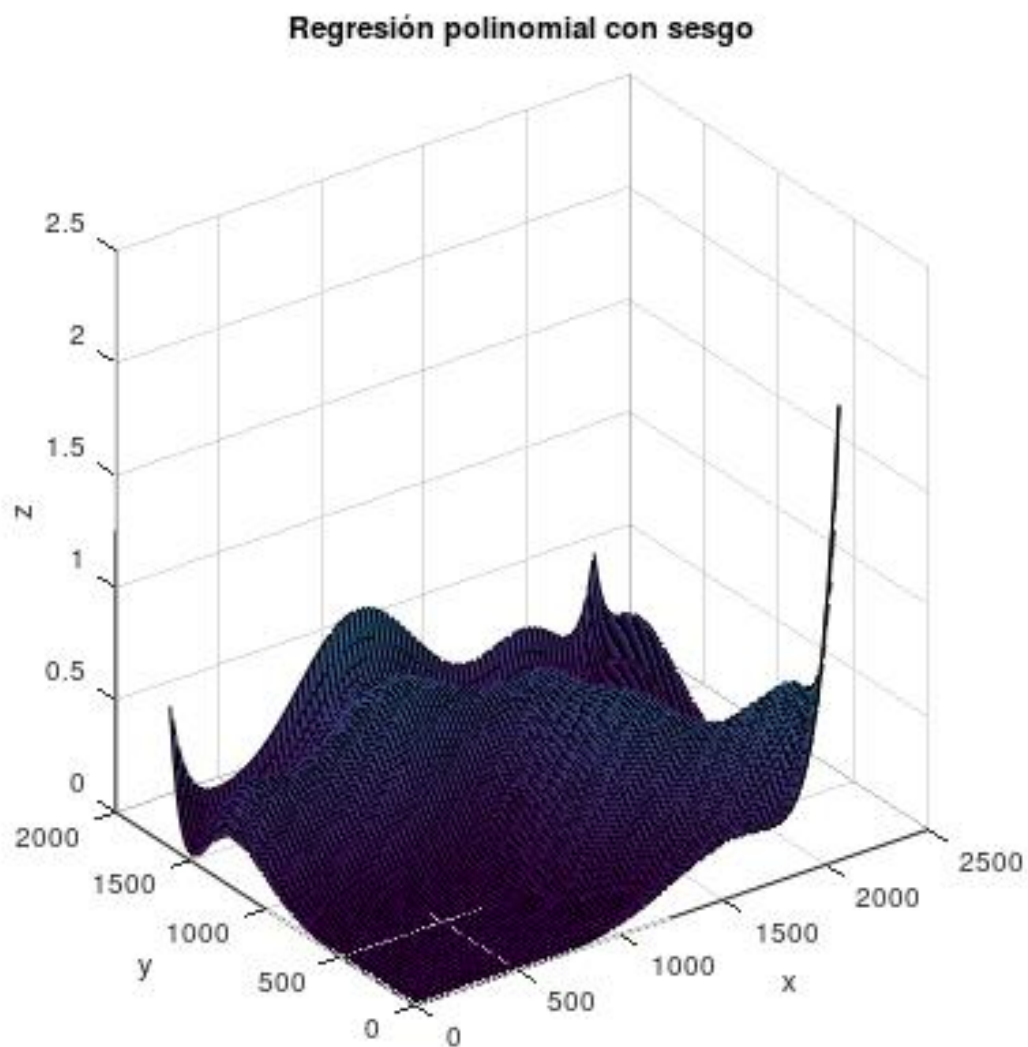


Figura 4: Regresión polinomial con sesgo.

La figura 5 muestra el error en función del grado del polinomio utilizado para generar la aproximación. A partir de esta se observa que los grados con menor error están en el rango de 5 a 10. Para la obtención del error se realiza una sumatoria del valor absoluto de las diferencias entre los datos censados “rz” (ground truth) y los datos aproximados para dividirla entre la cantidad de elementos en la sumatoria. Así también para el caso ponderado localmente

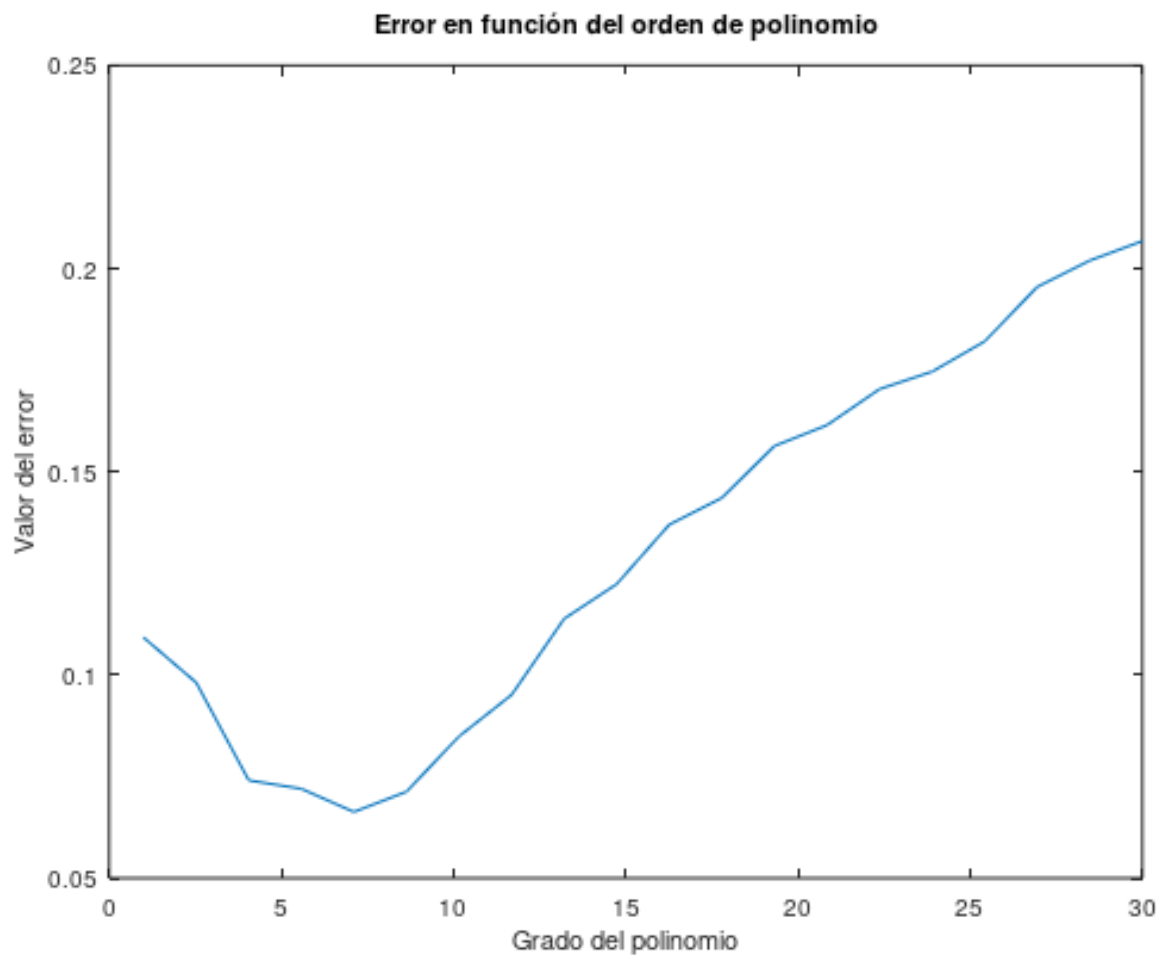


Figura 5: Error en función del orden del polinomio.

La figura 6 muestra la aproximación de la regresión polinomial con el menor grado de error.

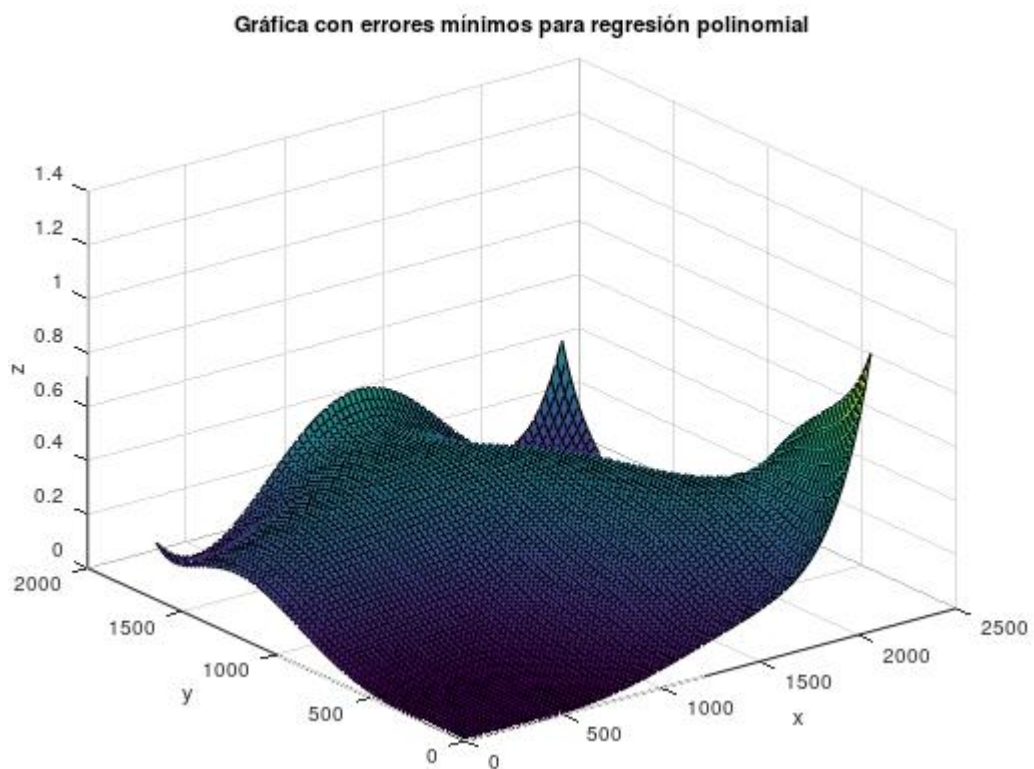


Figura 6: Regresión polinomial con menor grado de error.

La figura 7 muestra la aproximación utilizando una regresión ponderada localmente la cuál toma en cuenta la varianza “tau” de los datos para asignar “pesos” a los datos de entrenamiento a la hora de generar los parámetros theta. En el script “regressall” se define el parámetro tau deseado.

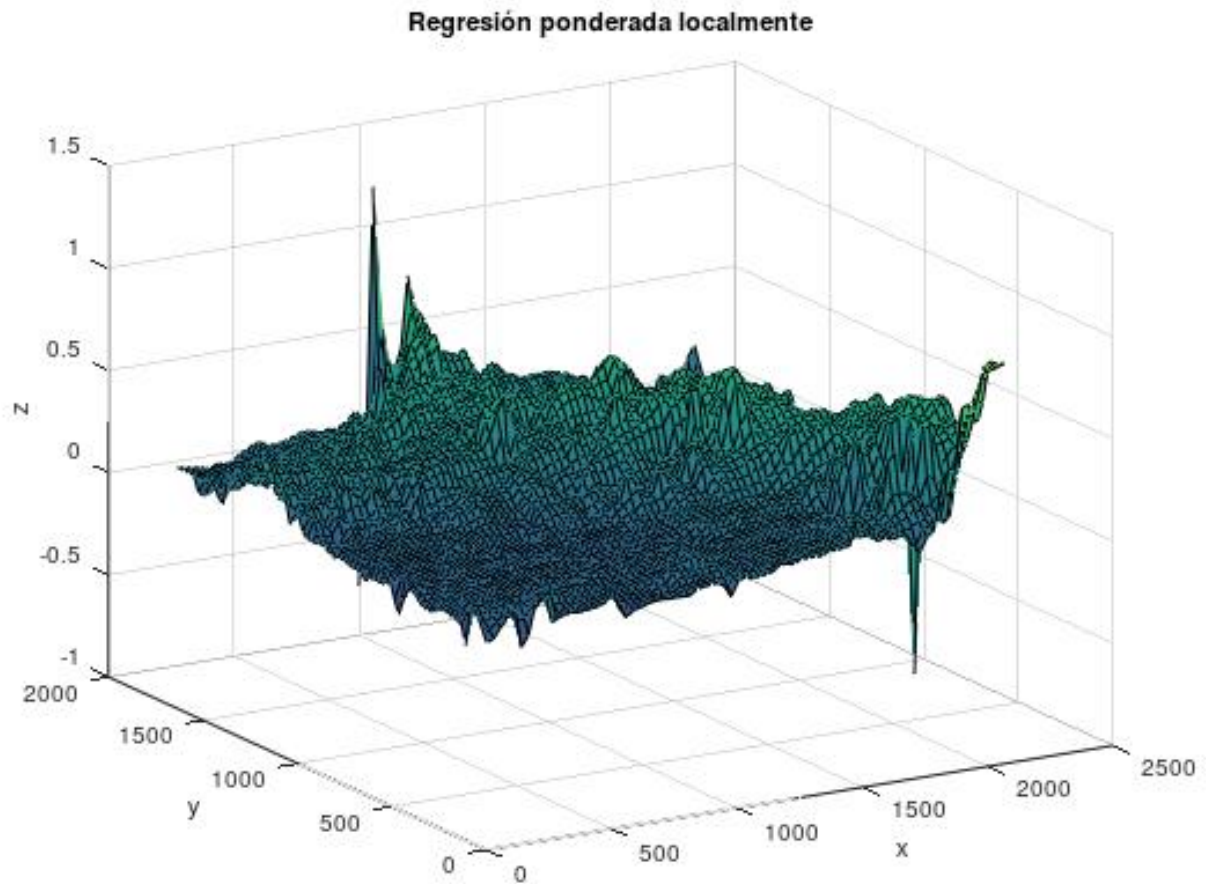


Figura 7: Regresión ponderada localmente.

La figura 8 muestra el error en función del valor de tau utilizado para generar la aproximación. A partir de esta se observa que los valores que producen menor error están en el rango de 15 a 20.

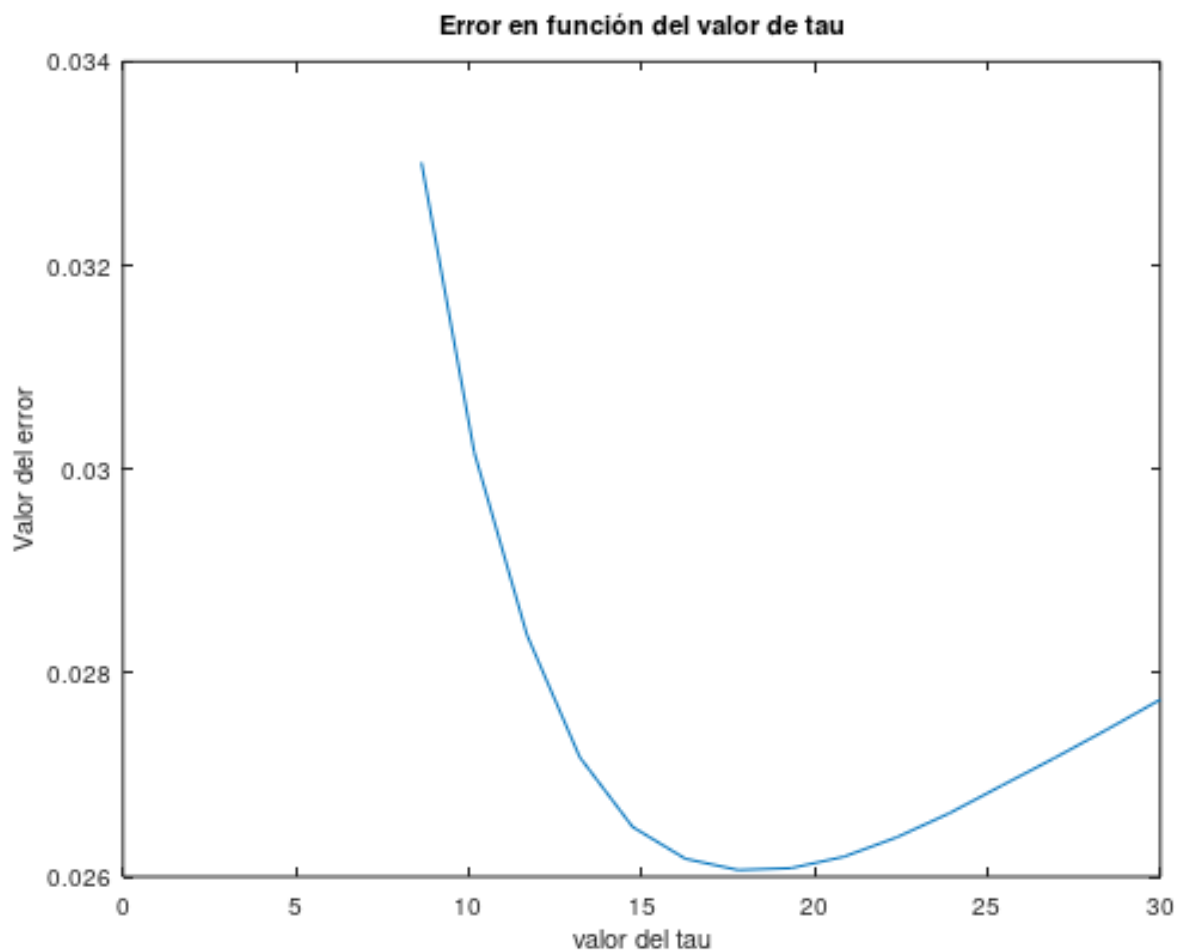


Figura 8: Error en función del valor tau.

La figura 9 muestra la aproximación de la regresión ponderada localmente con el menor grado de error.

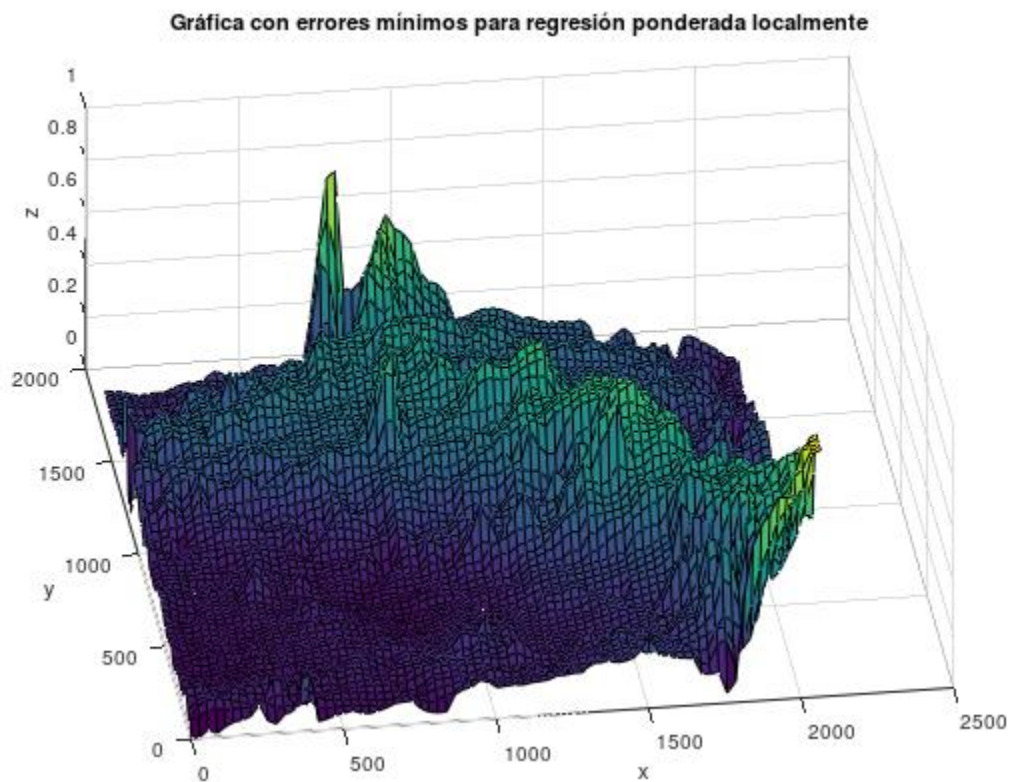


Figura 9: Errores mínimos para regresión ponderada localmente.

Conclusión: a partir de las gráficas de error en función del grado del polinomio en el caso polinomial y en función del valor de tau en el caso de la regresión ponderada localmente se observa que esta última fue el mejor método para aproximar la superficie del fondo marino.