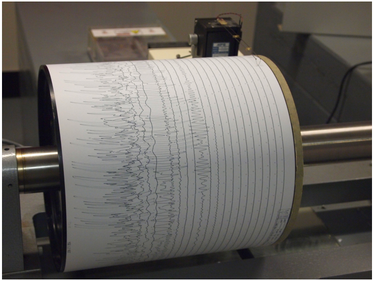
Regresión Lineal: Algunos Aspectos Prácticos

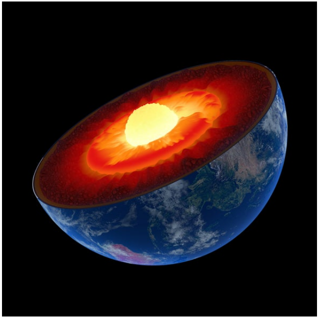
Mario I. Caicedo

3 de septiembre de 2021

Regresión≈Problema Inverso

$$\boldsymbol{d}^{(obs)} = \boldsymbol{F}(\boldsymbol{m})$$

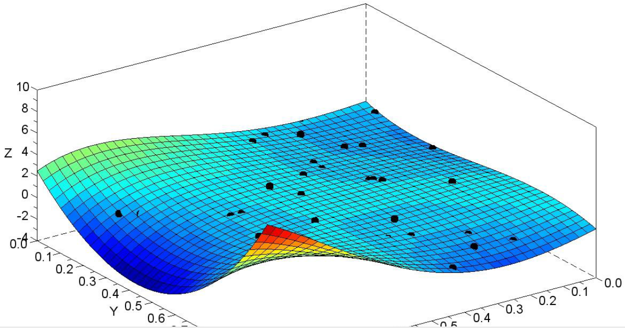




Regresión Lineal

$$\boxed{\boldsymbol{d} = \boldsymbol{F\Theta}}$$

$$\boxed{z_i = z_0 + a_i\, x_i + b_i y_i + a_{ij} x_i y_j + [\dots] + \epsilon_i}$$



Regresión Lineal

- El problema de **aprendizaje** consiste en encontrar (calcular) el vector de parámetros Θ
- La estructura de la soución comienza por la función de costo de mínimos cuadrados

$$J(\boldsymbol{m}) = (\boldsymbol{d}^{(obs)} - \boldsymbol{F}\boldsymbol{m})^T(\boldsymbol{d}^{(obs)} - \boldsymbol{F}\boldsymbol{m})$$

- El aprendizaje consiste en se reduce a minimizar la función de p'erdida con respecto a lo parámetos $(\boldsymbol{m})$.
- Solución analítica

$$\boxed{\boldsymbol{m} = (\boldsymbol{F}^T\boldsymbol{F})^{-1}\,\boldsymbol{F}^T\boldsymbol{d}^{(obs)}}$$

Regresión Lineal (Polinomio de segundo orden de una variable)

$$y = \theta_0 + \theta_1 x + \theta_2 x^2 \Longleftrightarrow \boldsymbol{y} = \boldsymbol{F\Theta}$$

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{pmatrix}, \qquad \boldsymbol{F} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ \vdots & \vdots & \vdots \\ 1 & x_N & x_N^2 \end{pmatrix}, \qquad \boldsymbol{\Theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{pmatrix}$$

Regresión Lineal Multivariable (Plano)

$$z = \theta_0 + \theta_1 x + \theta_2 y_1 \iff \boldsymbol{Z} = \boldsymbol{F\Theta}$$

$$\boldsymbol{Z} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_N \end{pmatrix}, \qquad \boldsymbol{F} = \begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \\ \vdots & \vdots & \vdots \\ 1 & x_N & y_N \end{pmatrix}, \qquad \boldsymbol{\Theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{pmatrix}$$

Regresión Lineal Multivariable (sup. de sgundo orden) I

$$z = \theta_0 + \theta_1 x + \theta_2 y_1 + \theta_3 xy + \theta_4 x^2 + \theta_5 y^2 \Longleftrightarrow \boldsymbol{Z} = \boldsymbol{F}\boldsymbol{\Theta}$$

$$\boldsymbol{Z} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_N \end{pmatrix}, \qquad \boldsymbol{F} = \begin{pmatrix} 1 & x_1 & y_1 & x_1 y_1 & x_1^2 & y_1^2 \\ 1 & x_2 & y_2 & x_2 y_2 & x_2^2 & y_2^2 \\ 1 & x_3 & y_3 & x_3 y_3 & x_3^2 & y_3^2 \\ \vdots & \vdots & \vdots & & & \\ 1 & x_N & y_N & x_N y_N & x_N^2 & y_N^2 \end{pmatrix}$$

$$\boldsymbol{\Theta} = \begin{pmatrix} \theta_0 & \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 \end{pmatrix}^T$$

Regresión Lineal (sup. de segundo orden) II

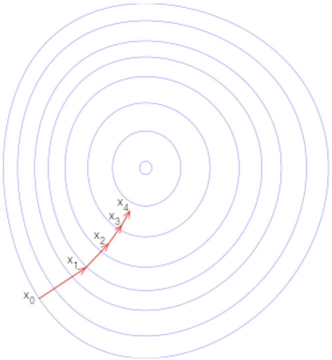
$$\boldsymbol{F}^T = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_N \\ y_1 & y_2 & \dots & y_N \\ x_1 y_1 & x_2 y_2 & \dots & x_N y_N \\ x_1^2 & x_2^2 & \dots & x_N^2 \\ y_1^2 & y_2^2 & \dots & y_N^2 \end{pmatrix}$$

Regresión Lineal (sup. de segundo orden) II

$$\boldsymbol{F}^T\boldsymbol{F} = \begin{pmatrix} N & \sum x_i & \sum y_i & \sum x_i y_i & \sum x_i^2 & \sum y_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i y_i & \sum x_i^2 y_i & \sum x_i^3 & \sum x_i y_i^2 \\ \sum y_i & \sum x_i y_i & \sum y_i^2 & \sum x_i y_i^2 & \sum y_i x_i^2 & \sum y_i^3 \\ \sum x_i y_i & \sum x_i^2 y_i & \sum x_i y_i^2 & \sum x_i^2 y_i^2 & \sum x_i^3 y_i & \sum x_i y_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^2 y_i & \sum x_i^3 y_i & \sum x_i^4 & \sum x_i^2 y_i^2 \\ \sum y_i^2 & \sum x_i y_i^2 & \sum y_i^3 & \sum x_i y_i^3 & \sum x_i^2 y_i^2 & \sum y_i^4 \end{pmatrix}$$

Note que esta matriz es simétrica y definida no negativa

Descenso por Gradiente



- La solución estándar consiste en minimizar la función de costo a través de la técnica de **Descenso por Gradiente**.

Descenso por Gradiente

- LOSS

$$J(\mathbf{\Theta}) = (\mathbf{d}^{(obs)} - \mathbf{F\Theta})^T(\mathbf{d}^{(obs)} - \mathbf{Fm})$$

- GRADIENTE

$$\boxed{\nabla_{\mathbf{\Theta}^T} J(\mathbf{\Theta}) = -\mathbf{F}^T(\mathbf{d}^{(obs)} - \mathbf{F\Theta})}$$

- ITERACIÓN

$$\mathbf{\Theta}^{(k+1)} = \mathbf{\Theta}^{(k)} - \mu \nabla_{\mathbf{\Theta}^T} J(\mathbf{\Theta}^{(k)})$$

- Veamos como funciona toda esta teoría