Regresión lineal univariada y multivariada. Parte 3

Pontificia Universidad Javeriana
Francisco Carlos Calderon Ph.D
2020

Objetivos

A partir del procedimiento hallado para una regresión de una variable, realizar el mismo procedimiento para varias variables.

Introducción al pre-procesamiento de las características y la iteración de variables de configuración (alpha)

Replanteando

$$h_\theta(x) = \theta_0 + \theta_1 x = y$$

$$\min_{\theta_0,\theta_1} \frac{1}{2n} \sum_{i=1}^{n} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right)^2 = \min_{\theta_0,\theta_1} \mathrm{J}(\theta_0, \theta_1)$$

$$h_\theta(\boldsymbol{x}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \cdots + \theta_m x_m$$

$$\min_{\theta_0,\theta_1,\ldots,\theta_m} \mathrm{J}(\theta_0, \theta_1, \ldots, \theta_m)$$

$$\min_{\theta_0,\theta_1,\ldots,\theta_m} \frac{1}{2n} \sum_{i=1}^{n} \left(h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right) - y^{(i)} \right)^2$$

Descenso del gradiente multivariado

Minimizar:

$$\min_{\theta_0,\theta_1} \mathrm{J}(\theta_0, \theta_1)$$

Repetir hasta encontrar convergencia:

$$\theta_j^{(k+1)} = \theta_j^{(k)} - \alpha \frac{\partial}{\partial \theta_j} \mathrm{J}(\theta_0, \theta_1)$$

Minimizar:

$$\min_{\theta_0,\theta_1,\dots,\theta_m} \mathrm{J}(\theta_0, \theta_1, \dots, \theta_m)$$

Repetir hasta encontrar convergencia:

$$\theta_j^{(k+1)} = \theta_j^{(k)} - \alpha \frac{\partial}{\partial \theta_j} \mathrm{J}(\theta_0, \theta_1, \dots, \theta_m)$$

Para el caso multivariado:

$$h_\theta(\boldsymbol{x}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \cdots + \theta_m x_m$$

$$\min_{\theta_0,\theta_1,\ldots,\theta_m} \frac{1}{2n} \sum_{i=1}^{n} \left(h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right) - y^{(i)}\right)^2$$

$$\theta_0^{(k+1)} = \theta_0^{(k)} - \alpha \frac{\partial}{\partial \theta_0} \mathrm{J}(\theta_0^{(k)}, \theta_1^{(k)}, \ldots, \theta_m^{(k)})$$

$$\theta_1^{(k+1)} = \theta_1^{(k)} - \alpha \frac{\partial}{\partial \theta_1} \mathrm{J}(\theta_0^{(k)}, \theta_1^{(k)}, \ldots, \theta_m^{(k)})$$

$$\theta_2^{(k+1)} = \theta_2^{(k)} - \alpha \frac{\partial}{\partial \theta_2} \mathrm{J}\left(\theta_0^{(k)}, \theta_1^{(k)}, \ldots, \theta_m^{(k)}\right)$$

.

.

.

Resolver ustedes mismos!

Para el caso multivariado:

$$h_\theta(\boldsymbol{x}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \cdots + \theta_m x_m$$

$$\min_{\theta_0,\theta_1,\ldots,\theta_m} \frac{1}{2n}\sum_{i=1}^{n}\left(h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right) - y^{(i)}\right)^2$$

$$\theta_0^{(k+1)} = \theta_0^{(k)} - \alpha\frac{1}{n}\sum_{i=1}^{n}\left(h_{\theta^k}\left(x^{(i)}\right) - y^{(i)}\right)$$

$$\theta_1^{(k+1)} = \theta_1^{(k)} - \alpha\frac{1}{n}\sum_{i=1}^{n}\left(h_{\boldsymbol{\theta}^k}\left(x^{(i)}\right) - y^{(i)}\right)\ x_1^{(i)}$$

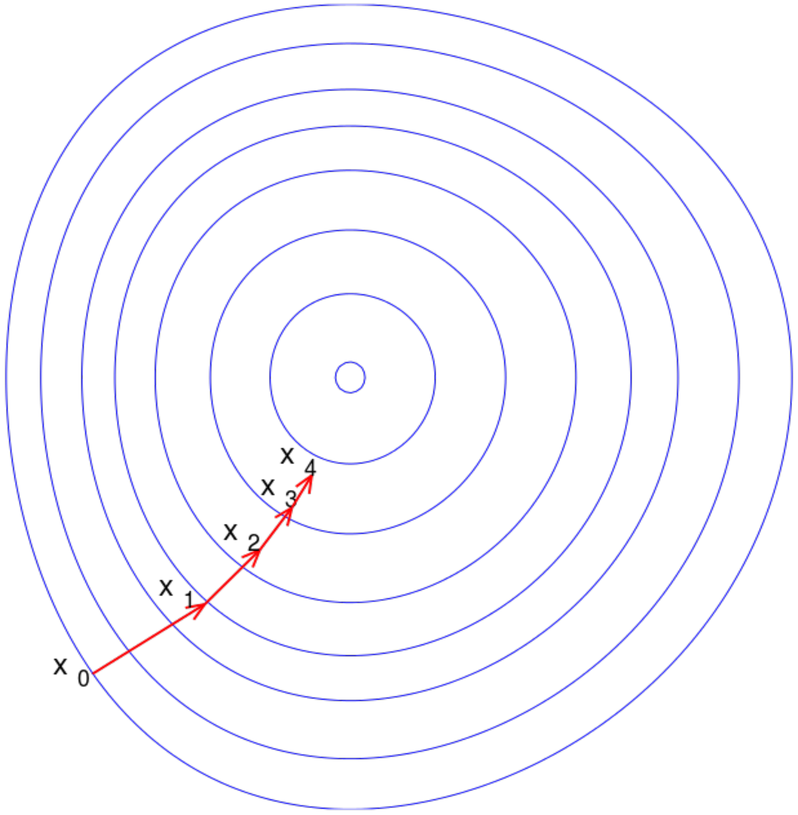
$$\theta_2^{(k+1)} = \theta_2^{(k)} - \alpha\frac{1}{n}\sum_{i=1}^{n}\left(h_{\boldsymbol{\theta}^k}\left(x^{(i)}\right) - y^{(i)}\right) x_2^{(i)}$$

$$\theta_m^{(k+1)} = \theta_m^{(k)} - \alpha\frac{1}{n}\sum_{i=1}^{n}\left(h_{\boldsymbol{\theta}^k}\left(x^{(i)}\right) - y^{(i)}\right) x_m^{(i)}$$

Pre procesamiento de características

Ya teniendo resuelta la regresión lineal

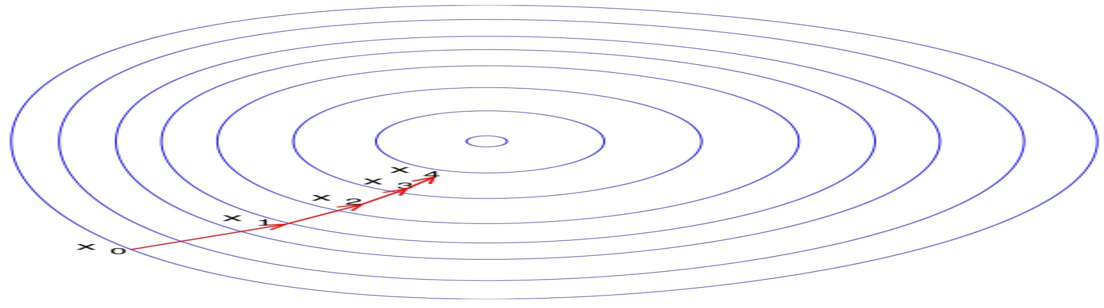
Caso ideal



Pre procesamiento de características

IDEA:

Pre procesar para mejorar la convergencia de los métodos de optimización.

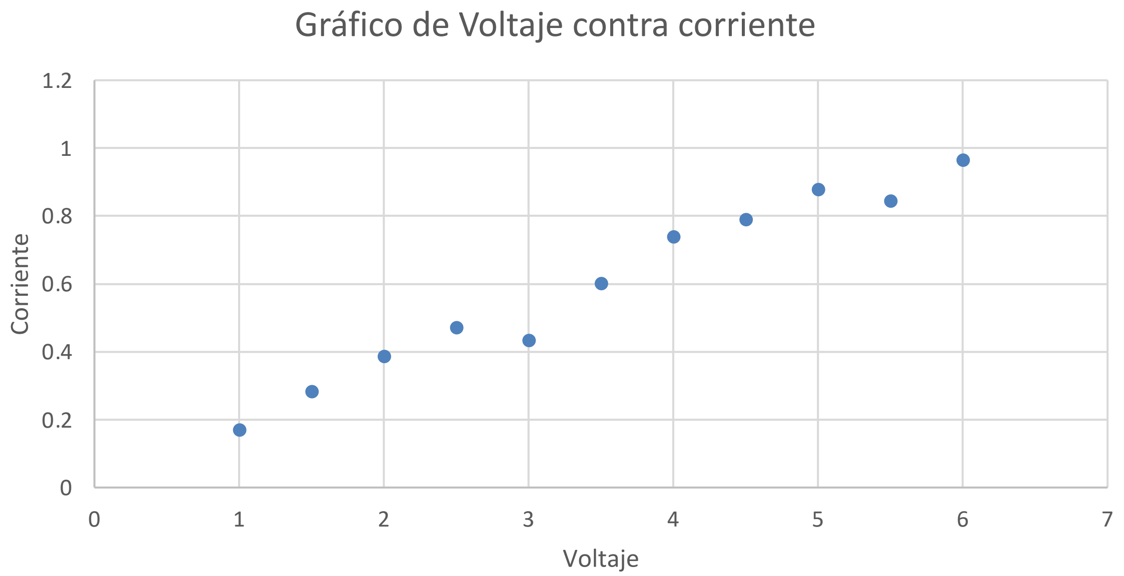


Ambos gradientes son afectados por el mismo alpha

- O uso alphas diferentes, -> no tan buena idea!
- O hago que mis características "x" sean similares. -> al parecer es una mejor idea!
  - Con esto también logro que mis x sean todos igual de importantes.

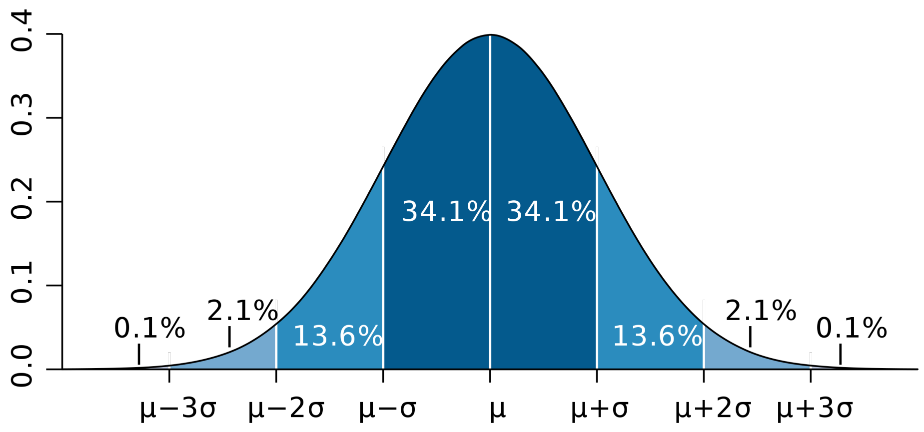
Normalización de características.

La normalización se hace necesaria ya que al tener características en intervalos similares se facilita la convergencia de los algoritmos de optimización.



j=	Voltaje	Corriente
0	1	0.16961027
1	1.5	0.28339581
2	2	0.38635874
3	2.5	0.47022787
4	3	0.43328129
5	3.5	0.60026765
6	4	0.73833898
7	4.5	0.79031502
8	5	0.87746427
9	5.5	0.84356446
10	6	0.96443892
Promedio	3.5	0.59611484
Desviación	1.58113883	0.25227499

Cómo normalizar?



A partir de la estadística si se supone una distribución

$$x' = \frac{x - \bar{x}}{\sigma}$$

Tomando el máximo y el mínimo "min-max"

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

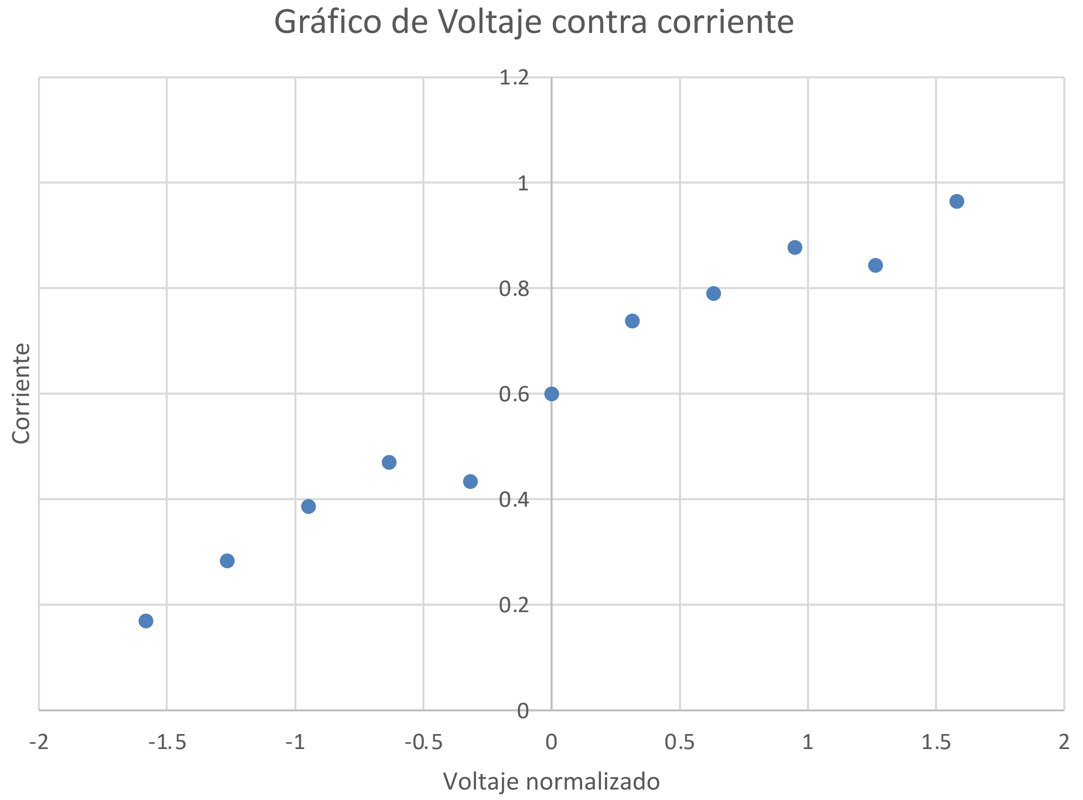
Centrando en la media

$$x' = \frac{x - \text{average}(x)}{\max(x) - \min(x)}$$

https://en.wikipedia.org/wiki/Feature_scaling

https://www.youtube.com/watch?v=9ONRMymR2Eg

Normalización de características.

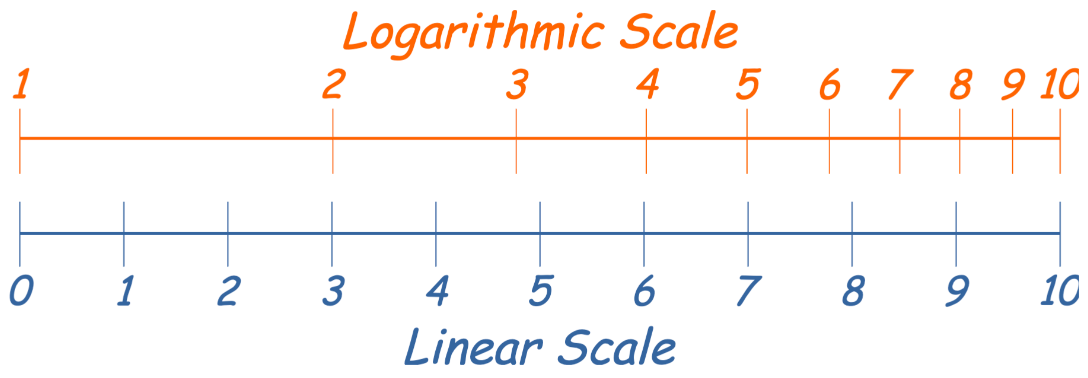


Voltaje Normalizado	Corriente
-1.58113883	0.16961027
-1.264911064	0.28339581
-0.948683298	0.38635874
-0.632455532	0.47022787
-0.316227766	0.43328129
0	0.60026765
0.316227766	0.73833898
0.632455532	0.79031502
0.948683298	0.87746427
1.264911064	0.84356446
1.58113883	0.96443892

Para este caso normalizamos restando el promedio de las muestras y dividiendo la resta por una desviación estándar

Promedio	3.5	0.59611484
Desviación	1.58113883	0.25227499

Selección del alpha



En la práctica es común tener que iterar entre diferentes alphas si no se logra una convergencia rápida o diverge.

Facultad de ingeniería
Depto de electrónica
Pontificia Universidad JAVERIANA Bogotá

Ejercicio en clase:

1. Normalizar x.
2. Encontrar el mínimo para J y verificar.
3. Encontrar un buen valor para alpha
4. Desnormalizar datos
5. Predecir usando nuestro regresor

6. Hacer caso multivariado

Probar con nuevos datos multivariados:

Ver excel adjunto hojas 3 y 4