

## Tema 4 : Regresión Multivariable

En este tema se describe el procedimiento para realizar una regresión a partir de un modelo que depende de múltiples variables. Para ello, considere los valores de la [Tabla 1](#). Aquí se listan trece muestras ( $m = 13$ ), las cuales están representadas como un conjunto  $\mathbf{x}$  que contiene tres características ( $\mathbf{x} = x_1, x_2, x_3$ ) y un objetivo ( $T$ ). Se utiliza la letra  $n$  para indicar el número de características, así, en este ejemplo  $n = 3$ . Al igual que en regresión de una sola variable, el propósito en regresión multivariable es encontrar un modelo  $y(x)$  (una función matemática) que relacione los valores de las características con el objetivo.

Table 1: Conjunto de datos empleados como referencia para explicar la regresión multivariable.

Muestra	$x_1$	$x_2$	$x_3$	$T$
1	1.74	5.3	10.8	25.5
2	6.32	5.42	9.4	31.2
3	6.22	8.41	7.2	25.9
4	10.52	4.63	8.5	38.4
5	1.19	11.6	9.4	18.4
6	1.22	5.85	9.9	26.7
7	4.1	6.62	8.0	26.4
8	6.32	8.72	9.1	25.9
9	4.08	4.42	8.7	32.0
10	4.15	7.6	9.2	25.2
11	10.15	4.83	9.4	39.7
12	1.72	3.12	7.6	35.7
13	1.7	5.3	8.2	26.5

De igual manera que en casos anteriores, se debe proponer el modelo (hipótesis) que pueda describir el comportamiento de las muestras. La idea inicial puede ser un polinomio de grado uno:

$$y(\mathbf{x}) = \omega_0 x_0 + \omega_1 x_1 + \omega_2 x_2 + \dots + \omega_n x_n \quad (1)$$

En esta notación, se utiliza el superíndice  $(i)$  para señalar a la  $i$ -ésima muestra ( $\forall i \in [1, m]$ ), de tal modo que:

$$y^{(i)}(\mathbf{x}^{(i)}) = \omega_0 x_0^{(i)} + \omega_1 x_1^{(i)} + \omega_2 x_2^{(i)} + \dots + \omega_n x_n^{(i)} = \sum_{j=0}^n \omega_j x_j^{(i)} \quad (2)$$

donde  $x_0^{(i)} = 1$  y el término  $\omega_0 x_0^{(i)}$  es usado para no restringir a la función a cruzar por el origen cuando las características tienen un valor cero. Una vez propuesta la hipótesis, se establece una función de costo de la cual hay que encontrar sus derivadas con respecto a cada coeficiente  $\omega$  afín de poder implementar el algoritmo de búsqueda GD. Como referencia se adopta el Error Medio Cuadrático, tal que:

$$\frac{\partial C(\mathbf{w})}{\partial w_j} = \frac{\partial}{\partial w_j} \frac{1}{2m} \sum_{i=1}^m (y^{(i)}(\mathbf{x}^{(i)}) - T^{(i)})^2 = \frac{1}{m} \sum_{i=1}^m (y^{(i)}(\mathbf{x}^{(i)}) - T^{(i)}) \frac{\partial}{\partial w_j} (y^{(i)}(\mathbf{x}^{(i)}) - T^{(i)}) \quad (3)$$

y

$$\frac{\partial}{\partial w_j} (y^{(i)}(\mathbf{x}^{(i)}) - T^{(i)}) = \frac{\partial}{\partial \omega_j} \left( \left( \omega_0 x_0 + \sum_{j=1}^n \omega_j x_j^{(i)} \right) - T^{(i)} \right) = x_j^{(i)} \quad (4)$$

por lo que las derivadas resultantes quedan como:

$$\frac{\partial C(\mathbf{w})}{\partial w_j} = \frac{1}{m} \sum_{i=1}^m (y^{(i)}(\mathbf{x}^{(i)}) - T^{(i)})x_j^{(i)} \quad (5)$$

A partir de este punto es posible implementar el algoritmo de búsqueda. Sin embargo, se puede considerar la opción de normalizar las características previo a aplicar el GD. En la [Tabla 2](#) se muestran los datos de la [Tabla 1](#) después de aplicar una normalización lineal para el rango  $[-1, 1]$ . Nótese como la normalización es aplicada a cada característica de manera individual, es decir, se realizaron tres normalizaciones para obtener un nuevo conjunto normalizado  $\mathbf{x}_{\text{norm}} = x_{\text{norm}_1}, x_{\text{norm}_2}, x_{\text{norm}_3}$  a partir de la normalización de  $x_1, x_2$  y  $x_3$  respectivamente.

Table 2: Resultado del proceso de normalización lineal a las características mostradas en la [Tabla 1](#)

Muestra	$x_{\text{norm}_1}$	$x_{\text{norm}_2}$	$x_{\text{norm}_3}$	$T$
1	-0.8821	-0.4858	1.	25.5
2	0.0996	-0.4575	0.2222	31.2
3	0.0782	0.2476	-1.	25.9
4	1.	-0.6438	-0.2777	38.4
5	-1.	1.	0.2222	18.4
6	-0.9935	-0.3561	0.5	26.7
7	-0.3762	-0.1745	-0.5555	26.4
8	0.0996	0.3207	0.0555	25.9
9	-0.3804	-0.6933	-0.1666	32.0
10	-0.3654	0.0566	0.1111	25.2
11	0.9206	-0.5966	0.2222	39.7
12	-0.8863	-1.	-0.7777	35.7
13	-0.8906	-0.4858	-0.4444	26.5

De ser empleada la normalización, la [Ecuación 2](#) se convierte en:

$$y^{(i)}(\mathbf{x}_{\text{norm}}^{(i)}) = \omega_0 x_{\text{norm}_0}^{(i)} + \omega_1 x_{\text{norm}_1}^{(i)} + \omega_2 x_{\text{norm}_2}^{(i)} + \dots + \omega_n x_{\text{norm}_n}^{(i)} = \sum_{j=0}^n \omega_j x_{\text{norm}_j}^{(i)} \quad (6)$$

mientras que la [Ecuación 5](#) queda expresada como:

$$\frac{\partial C(\mathbf{w})}{\partial w_j} = \frac{1}{m} \sum_{i=1}^m (y^{(i)}(\mathbf{x}_{\text{norm}}^{(i)}) - T^{(i)})x_{\text{norm}_j}^{(i)} \quad (7)$$

Con estas expresiones es posible aplicar el algoritmo GD obteniendo como resultado los coeficiente mostrados en la [Tabla 3](#) los cuales fueron el producto de la convergencia del GD ([Figura 1](#)). Note como al no utilizar normalización es necesario un valor de  $\alpha$  menor (para evitar la divergencia) y un número mayor de iteración en comparación a si utilizar la normalización. Asimismo, el valor de los coeficientes encontrados no es (ni debe ser) igual para los casos donde se usar normalización y el que no, esto es debido a que el normalizar los datos equivale a generar un nuevo conjunto de entrenamiento.

Finalmente, la [Tabla 4](#) muestra el los valores de predicción  $y$  para ambos casos; cuando la hipótesis recibe los datos sin normalizar ( $\mathbf{x}$ ) y cuando recibe los datos normalizados ( $\mathbf{x}_{\text{norm}}$ ). En esta tabla se puede apreciar que las predicciones para en ambos casos son aproximadamente iguales y el resultado del ajuste no es perfecto, puesto que se utiliza un polinomio de grado uno como hipótesis.

Con este resultado se podría sugerir que con un polinomio de grado uno es el mejor ajuste que se puede lograr. Dependiendo la aplicación y el propósito por el cual se desea realizar una regresión, se podría discutir acerca de como mejorar la hipótesis.

Table 3: Resultado del algoritmo GD.

Coeficientes	Valore encontrado por el GD	
	Sin Normalización	Con Normalización $[-1, 1]$
	$\alpha = 0.001$ Iteraciones = 500000	$\alpha = 0.5$ Iteraciones = 50
$\omega_0$	38.81241382	28.30836282
$\omega_1$	1.01961846	4.73327623
$\omega_2$	-1.85765944	-7.90642981
$\omega_3$	-0.30941042	-0.6240855

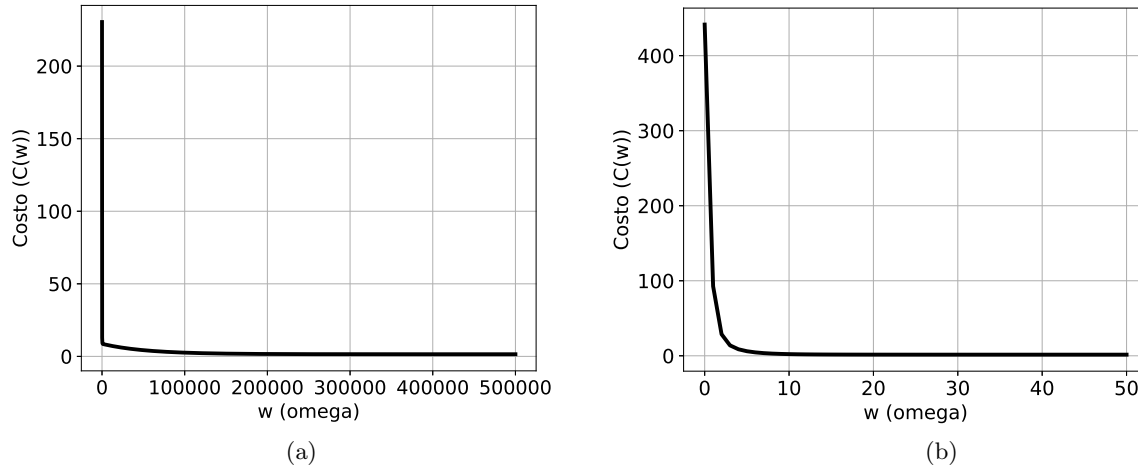


Figura 1: a) b) Costo obtenido para cada iteración.

Table 4: Resultados de la predicción  $y$  que toma cada una de las muestras como entrada.

Muestra	$y(\mathbf{x})$	$y(\mathbf{x}_{\text{norm}})$	$T$
1	27.3997	27.3503	25.5
2	32.2793	32.2590	31.2
3	27.3040	27.3447	25.9
4	38.3078	38.3056	38.4
5	15.5683	15.5297	18.4
6	26.1257	26.1092	26.7
7	28.2200	28.2542	26.4
8	26.2418	26.2092	25.9
9	32.0698	32.0937	32.0
10	26.0789	26.0614	25.2
11	37.2804	37.2451	39.7
12	32.4191	32.5049	35.7
13	28.1632	28.2113	26.5

## Ejercicios

1. Implementar un algoritmo de regresión multivariable para replicar el ejercicio descrito en este tema 4. Se debe realizar el ejercicio dos veces:

- (a) Sin emplear normalización
- (b) Empleando normalización

A partir de una base de datos “EjemploRM” ([Enlace](#)) realizar un algoritmo de regresión lineal multivariable que prediga correctamente la cantidad de CO<sub>2</sub> a partir de las características: “Volume” y “Weight”.