



Tecnológico de Monterrey

Módulo 2: Implementación de una técnica de aprendizaje máquina sin el uso de un framework.

Diego Armando Ulibarri Hernández

A01636875

Instituto Tecnológico y de Estudios Superiores de Monterrey

Inteligencia artificial avanzada para la ciencia de datos I (Gpo 101)

Análisis de resultados utilizando Regresión Lineal con gradiente descendiente implementada manualmente

Para este trabajo se seleccionó un dataset de bienes raíces con el cuál buscamos predecir el precio de las casas por metro cuadrado dependiendo de ciertas variables.

El primer paso para la implementación de nuestro modelo fue dividir nuestro dataset en 2 (train, test). esto con la finalidad de entrenar a nuestro modelo con el dataset de train que contiene el 80% de nuestros datos y una vez entrenado ver cómo predice los precios de dataset de test el cuál nunca antes ha visto. Además de dividir nuestro dataset también escalamos los datos para poder trabajar de una mejor manera con ellos.

Finalmente ya que hicimos la división de los datos y generamos nuestro modelo podemos realizar algunas pruebas variando nuestro learning rate y las épocas y de esta forma lograr obtener nuestro mejor coeficiente de determinación que en este caso y realizando varias pruebas fue **0.39**.

```
Coeficiente de determinación: 0.39456825313772026
learning rate: 0.01      épocas: 10000
  Valor deseado  Valor obtenido
0      0.527844      0.818102
1     -0.345957      0.519288
2     -0.997948     -0.811706
3     -1.112214     -0.767005
4      0.971467      0.685871
..      ...
78     -0.278742     -0.045567
79     -0.487110      0.397615
80     -1.105493     -0.585633
81      0.070779      0.805031
82      0.709326      0.903344

[83 rows x 2 columns]
```



```
Coeficiente de determinación: 0.16039876936317754
learning rate: 0.0001    épocas: 80000
  Valor deseado  Valor obtenido
0      0.527844      1.153378
1     -0.345957      0.918274
2     -0.997948     -0.472634
3     -1.112214     -0.928901
4      0.971467      0.903788
..      ...
78     -0.278742      0.062460
79     -0.487110      0.832066
80     -1.105493     -0.840937
81      0.070779      0.911831
82      0.709326      1.111441
```

Como conclusión el coeficiente de determinación obtenido nos indica que nuestro modelo en este caso particular no es tan confiable, sin embargo, puede ser que los datos que estamos utilizando tampoco sean los más adecuados para el modelo o que incluso si estos datos los tratamos adecuadamente tal vez lograríamos tener un mejor score.