



# Tecnológico de Monterrey

Inteligencia Artificial Avanzada para la Ciencia de Datos

**Modelo de Regresión lineal utilizando un framework**

Renata de Luna Flores

A01750484

Profesor: Jorge Ramírez Uresti

Implementando el desarrollo de un modelo de regresión lineal con la ayuda de un framework podemos calcular el score, los errores y además hacer predicciones.

En primer lugar, debemos dividir el dataframe en train y test para con uno poder entrenar el modelo y con otro probarlo. El tamaño de los dos debe estar correctamente estipulado, ya que, si es muy grande o pequeño, el modelo puede tener problemas como overfitting o de underfitting. En las predicciones realizadas y las diferentes formas de separar el data frame original, podemos ver que, al tener un test size muy grande o muy pequeño el error incrementa y el score (coeficiente de determinación de la predicción) es más bajo que al tener un test size común (como lo es 20-25%). Además de esto, al realizar diferentes corridas del modelo creado, el score y el error generado por el modelo varían, debido a que los datos para train y test cambian al aleatorio.

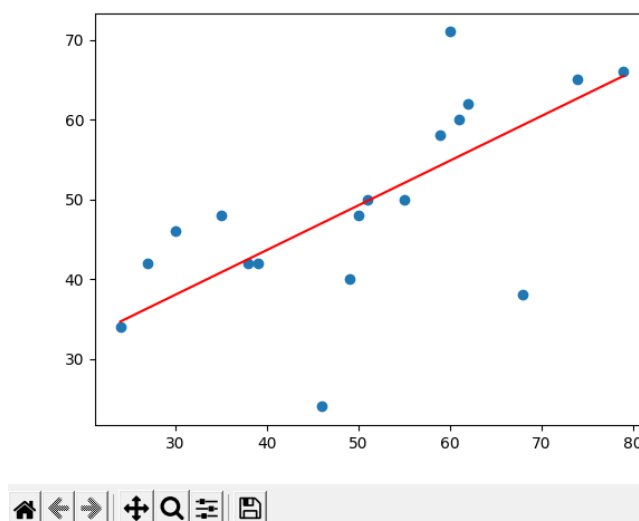
En el primer ejemplo utilizado podemos ver que el score del modelo no es malo, aunque podría mejorarse.

```
The model's score is: 0.8691573685051167

The predictions for the x test are:
      0
0  71.244587
1  44.290792
2  63.510114
3  25.175867
4  48.826734
Values for Y
24    75
18    46
22    52
21    26
16    52
Name: satisfaction, dtype: int64

The model's errors are:
MAE: 4.1944268071820145
MSE: 32.05121101098661
```

Figure 1



Podemos ver que el modelo tiene un buen score, lo que además se puede apreciar en la gráfica, ya que el sesgo no es muy grande.

Teniendo más datos de training, el score del modelo mejora, pero podríamos caer en overfitting o en underfitting, y que en realidad el modelo solo está aprendiendo los valores o no aprende totalmente.

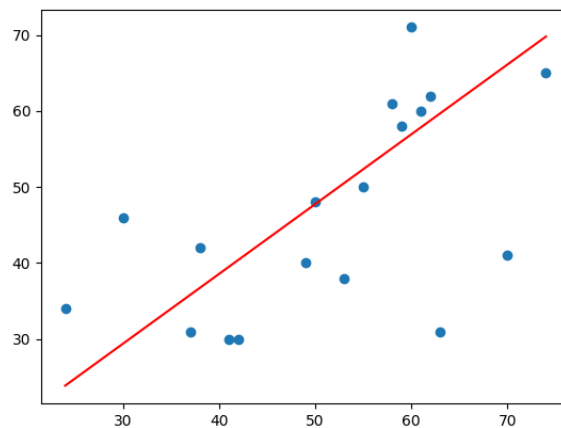
```
Example 2
Test size of 50%
The model's score is: 0.886508905510127
```

```
The predictions for the x test are:
```

```
0 82.650553
1 86.140475
2 42.206745
3 24.236108
4 66.131824
5 61.500288
6 82.547123
7 85.046121
8 89.953473
9 45.097774
10 63.908640
11 94.688439
12 81.556200
7 88
11 88
17 43
21 26
15 70
22 52
3 80
```

```
The model's errors are:
MAE: 6.183104007270213
MSE: 51.57704575526998
```

Figure 1



Teniendo solo el 50% del tamaño del data set, el score del modelo sigue siendo bueno. Esto también puede ser visto en la gráfica obteniendo un sesgo bajo.

Finalmente, teniendo solo 10% de los valores de los datos para hacer pruebas, el modelo tiene un score muy bajo, por lo que caemos en overfitting.

```

Example 3
Test size of 10%
The model's score is: -1.2037500433494062

The predictions for the x test are:
0  48.159728
1  68.829737
2  63.982384
0
0  48.159728
1  68.829737
2  63.982384
Values for Y
5    44
10   56
9    57
Name: satisfaction, dtype: int64

The model's errors are:
MAE:  7.990616349656601
MSE: 76.88639040130151

```

Al hacer pruebas solo con el 10% de los datos (y considerando que tenemos pocos datos), el score del modelo es inclusive negativo, por lo que podemos decir que no es una buena separación de los datos.

Además, podemos realizar predicciones con datos que introducimos de la edad y de la severidad de la enfermedad:

```

Making predictions with input data:
Introduce the sick person's age: 60
Introduce the sickness severity: 90
Satisfaction: 32.00139371334106

```

```

Making predictions with input data:
Introduce the sick person's age: 100
Introduce the sickness severity: 80
Satisfaction: -4.42946635850592

```

```

Making predictions with input data:
Introduce the sick person's age: 15
Introduce the sickness severity: 50
Satisfaction: 97.10284604161359

```

Con el uso de un framework no podemos cambiar diferentes parámetros para mejorar la precisión de este mismo. Por esto, debemos asignar un tamaño correcto al test size. Sin embargo, el score que tiene el modelo al ser implementado con un framework no es bajo y el error no es muy importante, por lo que podemos decir que tiene un buen accuracy.

Finalmente, con la implementación podemos realizar predicciones introduciendo la edad y la gravedad de la enfermedad, para predecir cuánto puede el paciente sanar.

