



Tecnológico de Monterrey

Campus Monterrey
Retroalimentación Módulo 2

Inteligencia Artificial Avanzada para la Ciencia de Datos
TC3006C Grupo 102

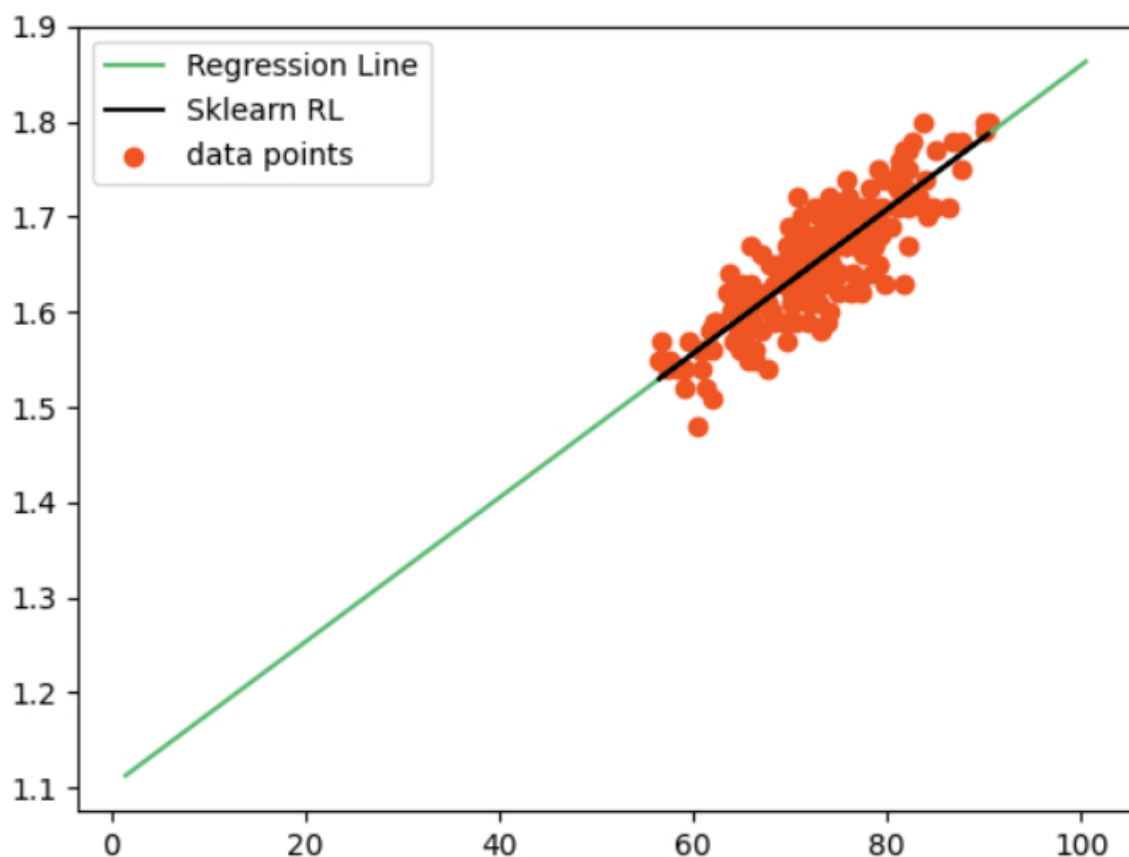
Alumno:

Omar Enrique González Uresti

a domingo 10 de septiembre del 2022

En la entrega 1 de éste módulo se implementó un modelo basado en el problema de Kaggle del Titanic, sin embargo, se desechó ya que la mayoría de las variables no son dependientes de las otras, exceptuando la de supervivencia. Fue por esto que se decidió cambiar a otro dataset; este de peso y estatura.

Gráfica con un modelo sin librerías (verde) y con Scikit Learn (negro). A simple vista no se ve diferencia. En la X se ve el peso, y la Y la estatura.

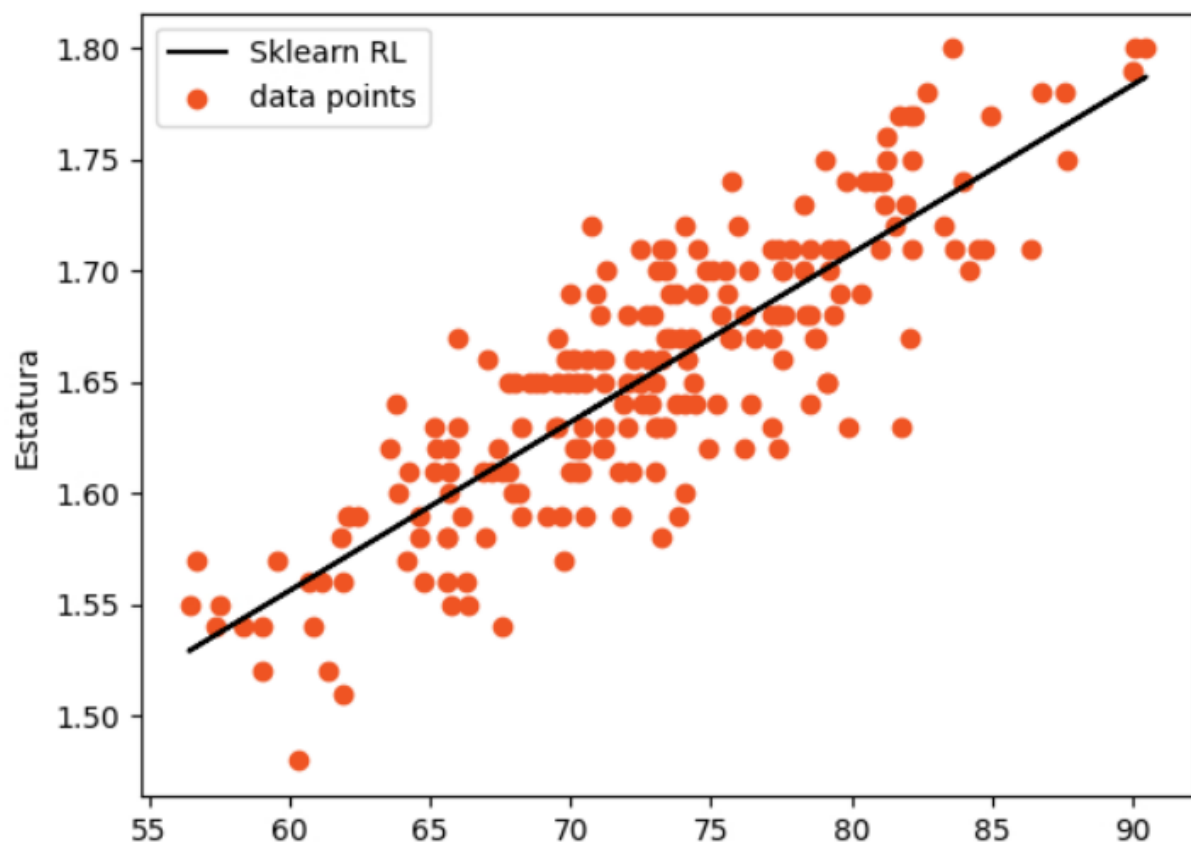


Viendo que la estatura empieza a partir de 1.40 aproximadamente, se intuye que entonces esta base de datos fue llenada con personas adultas.

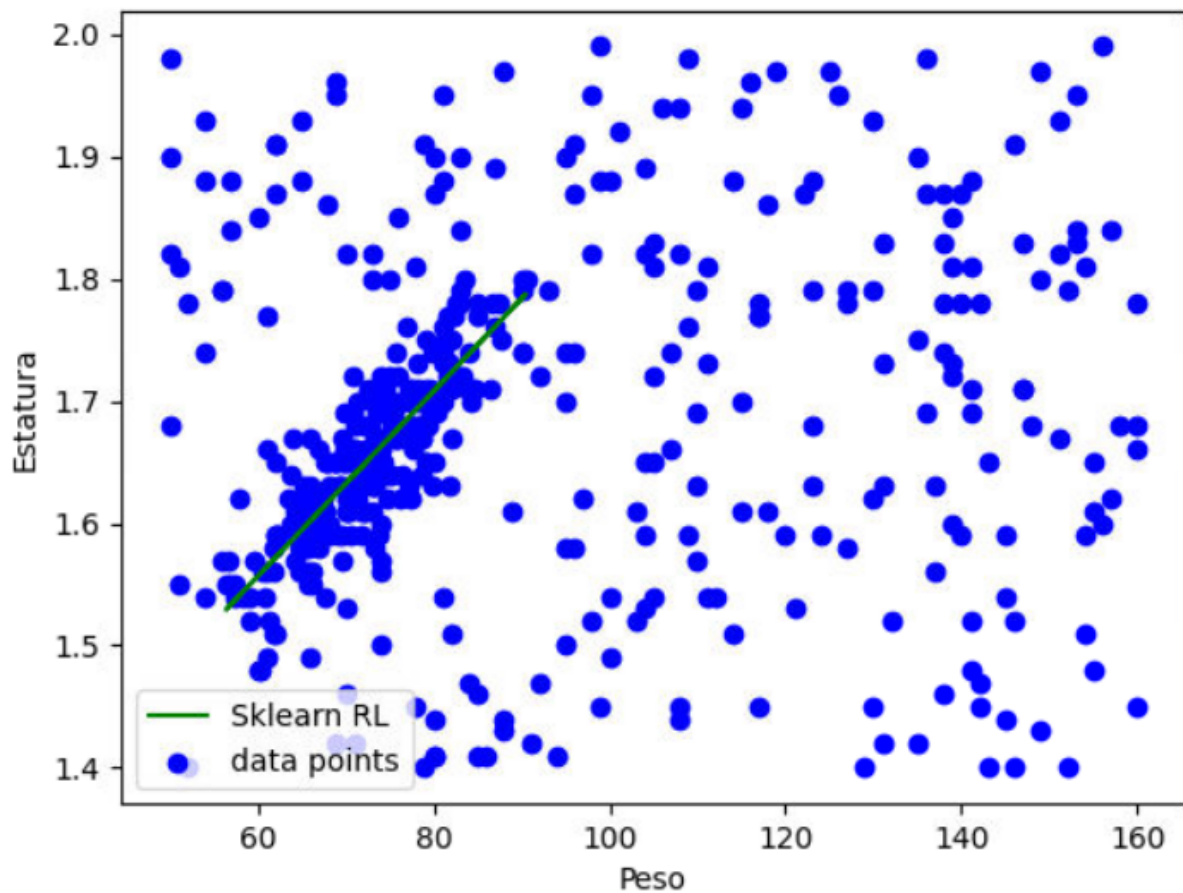
También viendo la gráfica, no existen personas que midan 0 o pesen 0, esto es lo lógico.

Rcuadrado de modelo original (r^2) y de Scikit Learn (r^2_{sl}). Se concluye que scikit learn es mejor modelo porque su R^2 es mejor.

```
r2 = 0.6395575198599801
r2sl = 0.7171291648571596
```



Gráfica de sólamente datos sacados por Scikit Learn Linear Regression usando un dataset pequeño, el cual servirá como Training file.



Prueba con un dataset con más datos, el cual funcionaría como validation, sin embargo, se nota que no se cumple con un modelo funcional, a causa de que muy probablemente el dataset fue hecho con datos randomizados en vez de lógicos.

Rcuadrado del modelo con el dataset de validation.

0.013988008581686473

Se concluye que el modelo es underfitt porque no cumple con una estimación precisa de los valores, además que el training no sirvió de mucho a la hora de ponerlo en validation.

Github del proyecto

<https://github.com/OmarUresti38/RetoModulo2.2>