

En este documento se evaluará el desempeño de ambos modelos, la implementación de linear regression sin y con framework y se escogerá uno de los 2 para realizar el siguiente análisis: diagnóstico y explicación del grado de sesgo, grado de varianza y del nivel de ajuste del modelo.

El código se encuentra en el siguiente repositorio público:

https://github.com/eprado99/framework_impl

Para ambas implementaciones se utilizó un data set compartido en el modulo de Estadística, el cual contiene las estaturas y el peso de Hombres y Mujeres.

Separamos nuestro data set en train y test utilizando la función de Sklearn `train_test_split` e hicimos únicamente uso de los datos del data set de los hombres.

Comenzaremos listando ambas implementaciones, la primera implementación es la de Regresión Lineal sin uso de un framework.

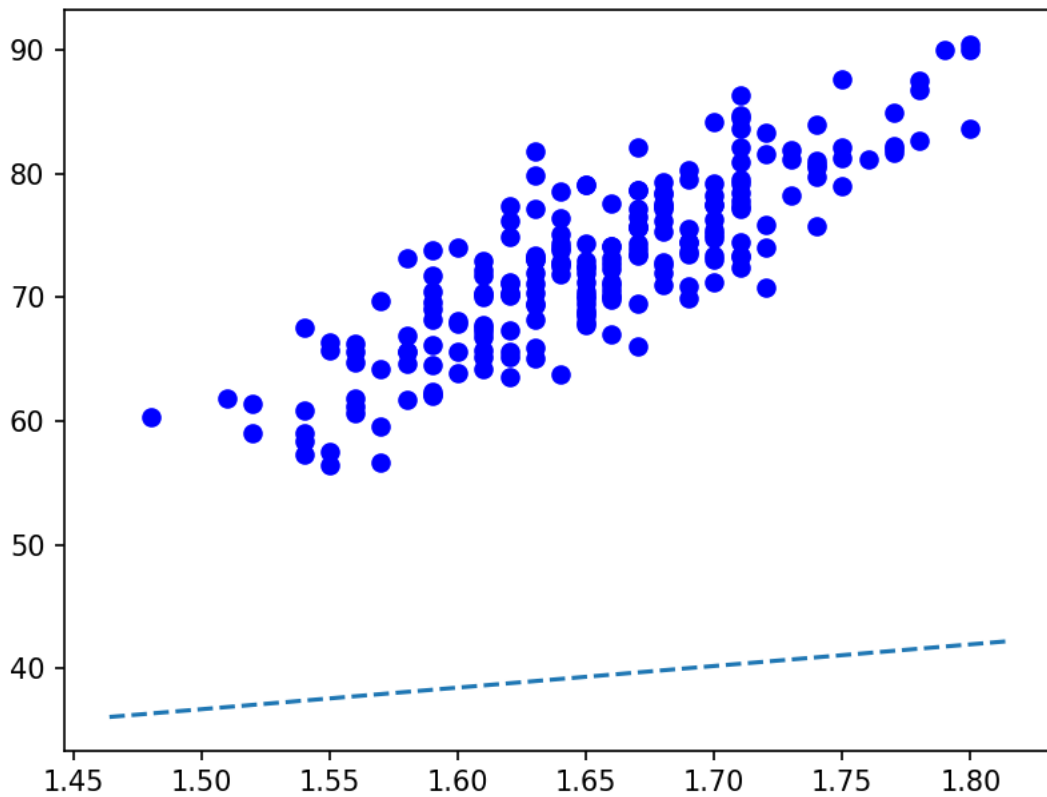
Regresión Lineal sin uso de un framework

Para esta entrega se investigo el algoritmo de optimización Gradient Descent y se utilizo en conjunto con la función de costo Mean Squared Error para medir el error.

Se utilizaron 2 derivadas para lograr la optimización, una con respecto a la slope y otra con respecto al intercept. Estas mismas moviéndose hacia el local mínima.

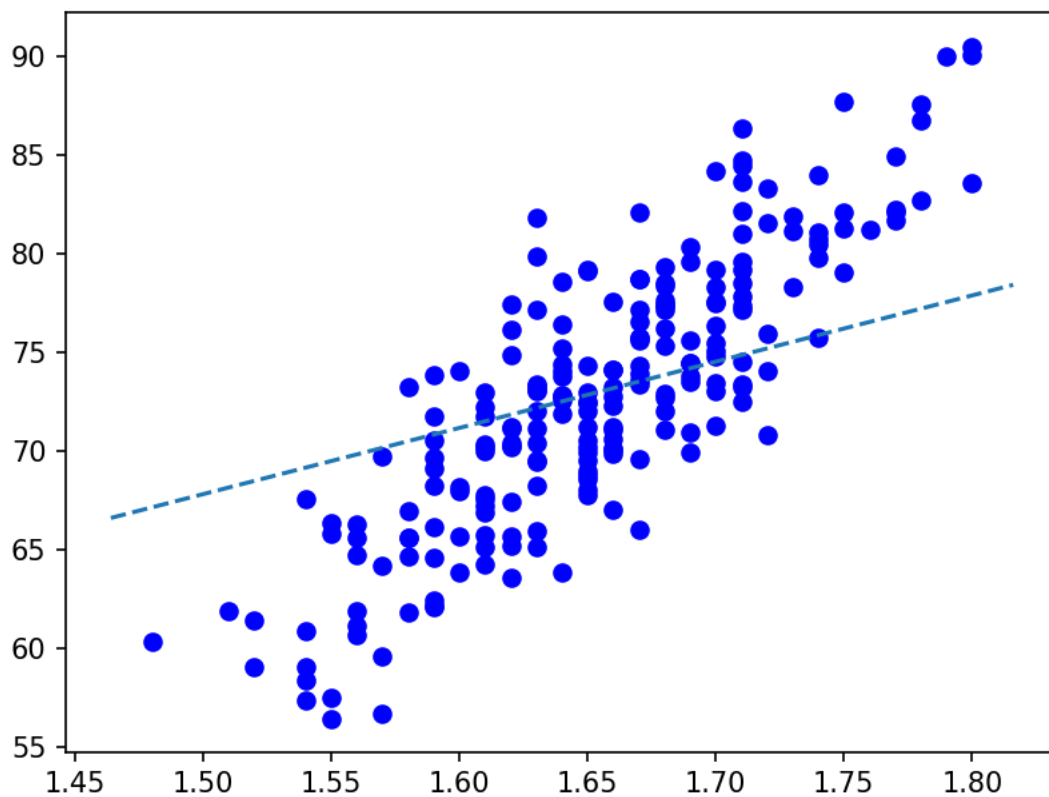
Se configuro con un Learning Rate (Alpha) de 0.01, este siendo el tamaño de los pasos tomados para llegar al local mínima.

Aquí una grafica del algoritmo corriendo por 10 épocas, lo cual probó ser un numero de épocas muy bajo por lo cual se ajustó.



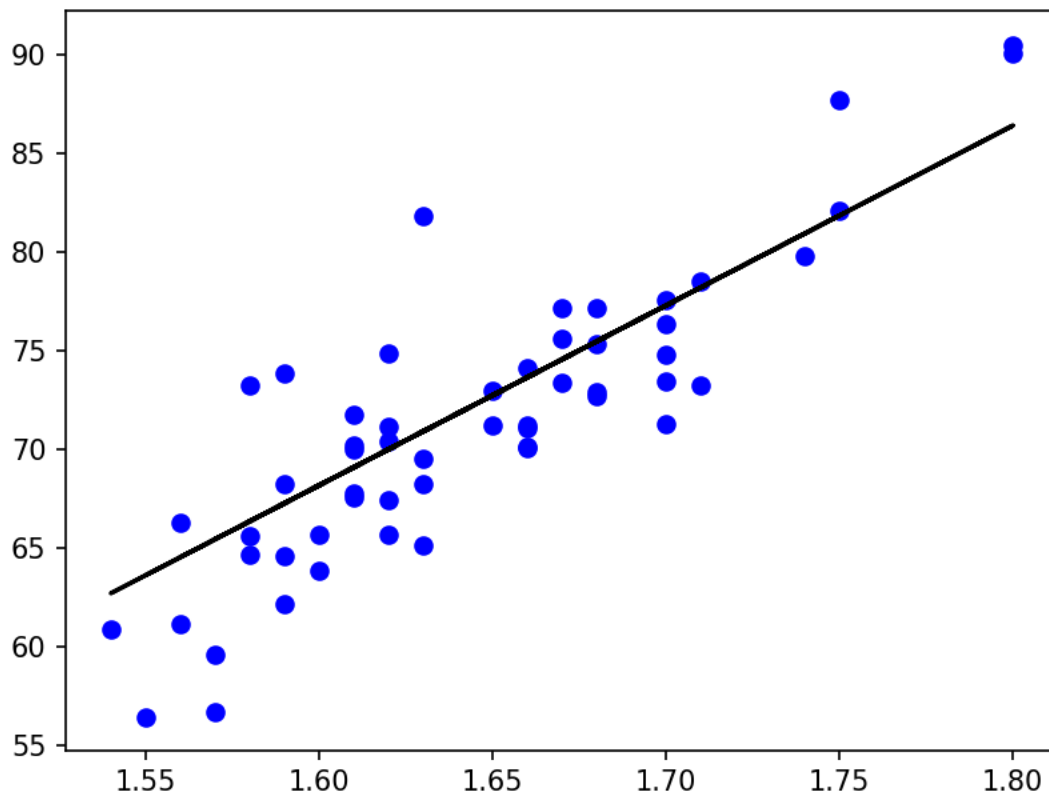
Se procedió a correr el algoritmo por 1000 épocas para ajustar la recta.

A continuación, podemos ver el resultado del algoritmo con nuestra línea ajustada al modelo, teniendo el eje Y como el peso y el eje X como la estatura.



Podemos visualizar en esta grafica que no se ajusta correctamente la recta a los datos, logra pasar por algunos de ellos, pero hay varios datos que se encuentran lejos de la recta, por lo cual hay mucha área de mejora para este algoritmo, pero para los usos de este algoritmo es algo aceptable. A continuación, hablaremos de la implementación de Regresión Lineal utilizando el framework de Sklearn para lograr hacer la comparativa entre ambos.

Para esta implementación utilizamos la función de LinearRegression importada de sklearn. La definimos con los parámetros por default y realizamos el entrenamiento de la misma manera que sin el uso del framework.



Aquí inmediatamente podemos ver como obtenemos una recta mucho mas acertada que pasa por la mayoría de los datos en la gráfica, en esta grafica se muestran menos datos sin embargo esto es porque solo se esta graficando contra los datos de prueba. Con poca configuración rápidamente nos dio una solución mas acertada que nuestra implementación lo cual indica que podemos mejorar muchísimo más nuestro algoritmo inicial.

Ahora para realizar el análisis sobre una de las 2 implementaciones, escogeremos la implementación de Regresión Lineal con el uso del framework de Sklearn.

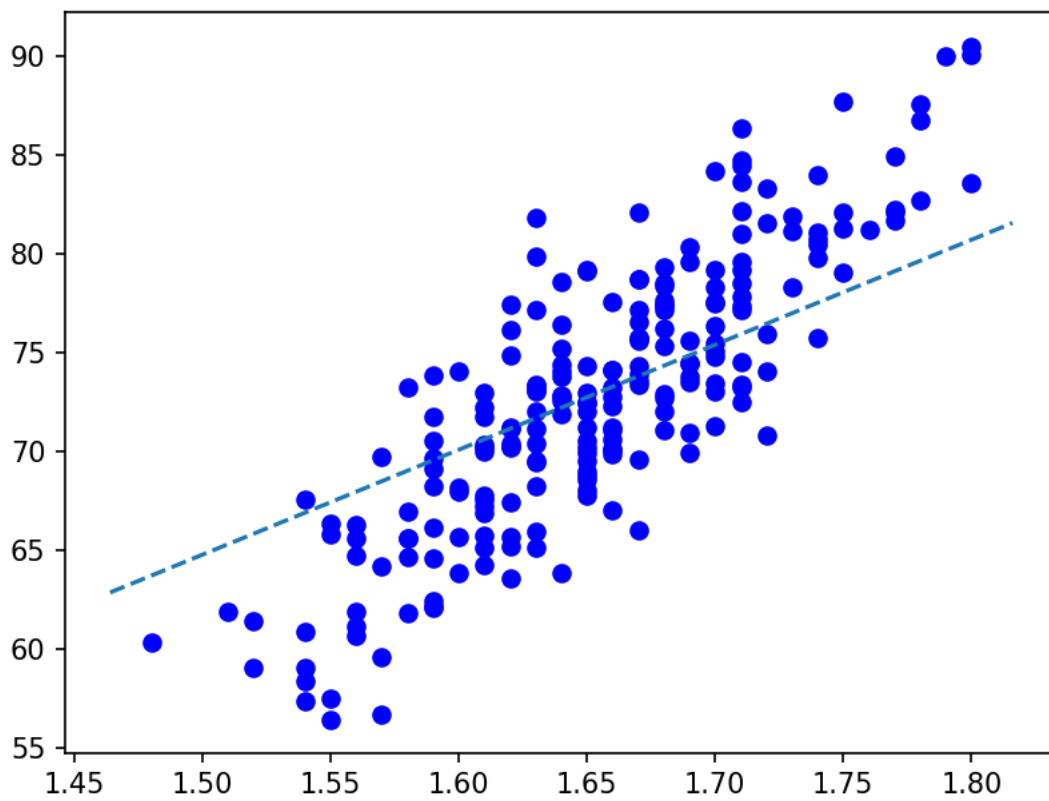
1 – En esta implementacion rápidamente podemos notar como tiene un grado de sesgo bajo al ajustarse a los datos mas comunes y evitar aquellos considerados outliers.

2 – El grado de varianza también es bajo ya que no suele variar mucho la recta y se encuentra sobre la mayoría de los datos por lo que se acomoda bastante bien.

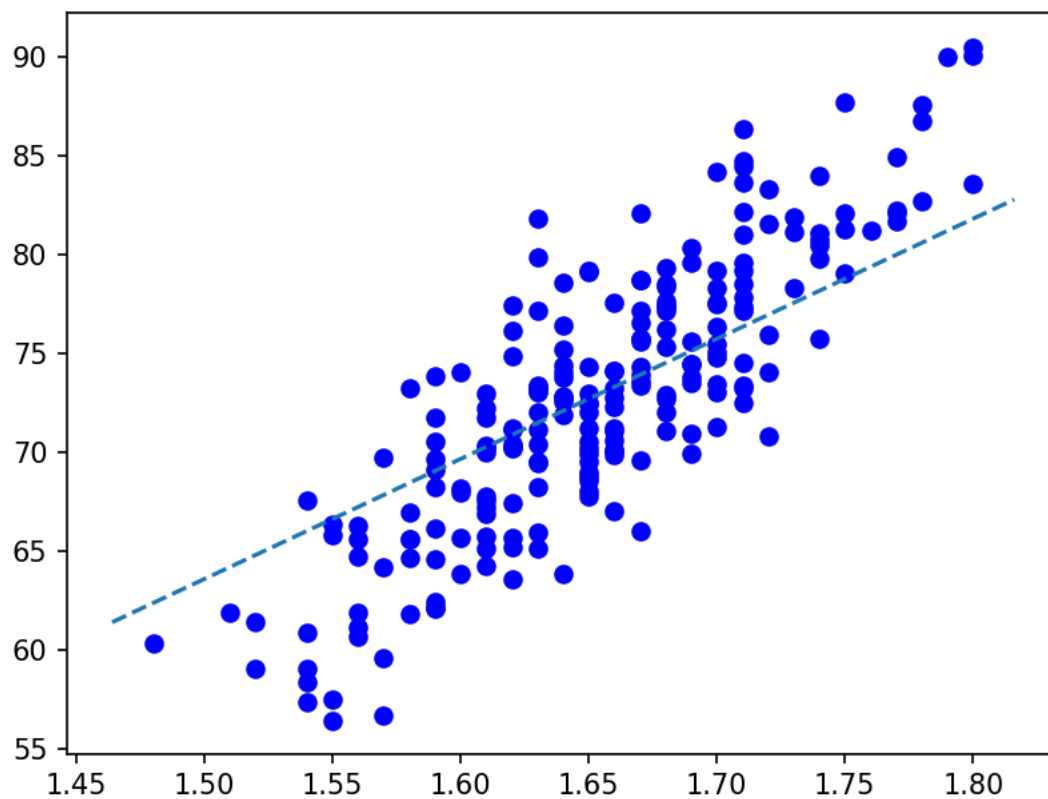
3 – El nivel del ajuste del modelo se considera como fit ya que se acomoda de manera correcta a los datos y se logran hacer predicciones con un bajo MSE.

Mejora del modelo sin el uso de framework:

Para lograr mejorar el modelo se procedió a ajustar la variable Alpha (Learning Rate) de 0.01 a 0.2 lo cual hizo que la recta se ajustara de mejor manera a los datos, es con esto que se logro mejorar el modelo al ajustar el parámetro



Ahora ajustando el número de épocas de 1000 a 1500 logro una mejoría algo significativa, por lo que la recta termina ajustándose mejor a los datos.



Por último, ajustando las épocas de 1500 a 3000 se logro obtener una recta la cual se ajusta muchísimo mejor que antes al modelo, entonces al ajustar nuestra Alpha y el numero de épocas en nuestro algoritmo sin el uso de framework, logramos acercarnos mucho mas a la recta que nos proporciona el algoritmo de Regresión Lineal de Sklearn.

