

## **Análisis y Reporte sobre el desempeño del modelo de regresión lineal**

**José Ángel García López**  
**A01275108**

### **Inteligencia artificial avanzada para la ciencia de datos**

#### **Modelo**

El modelo implementado para esta entrega es una regresión lineal la cual es una técnica fundamental en estadísticas y aprendizaje automático que se utiliza para modelar la relación entre una variable independiente y una variable dependiente. En este informe, presentamos una implementación práctica de una regresión lineal simple utilizando la plataforma de aprendizaje profundo TensorFlow.

#### **Set de datos**

El set de datos utilizados en este avance es el del iris el cual comúnmente se utiliza para problemas de clasificación, sin embargo quise tomar un enfoque diferente analizando la relación entre dos de sus componentes como lo es el ancho del pétalo y el largo del sépalo.

#### **Separación de datos**

Para este modelo se dividió en 60% para train, 20% para test y 20% para validación, realmente no hay una razón en específico para estos porcentajes y de hecho es algo que se modificará en el futuro para demostrar la forma en la que cambian los resultados.

#### **Resultados**

En una etapa inicial, se realizaron diferentes pruebas con los datos crudos, esto quiere decir que simplemente importamos los datos y tal y como venían se introdujeron al modelo dando un resultado relativamente bueno dada la naturaleza del modelo implementado, cabe destacar que se implementaron 100 epochs y se utilizó el método de optimización de Gradiente descendiente usando un “learning rate” de 0.03, los resultados se presentan a continuación:

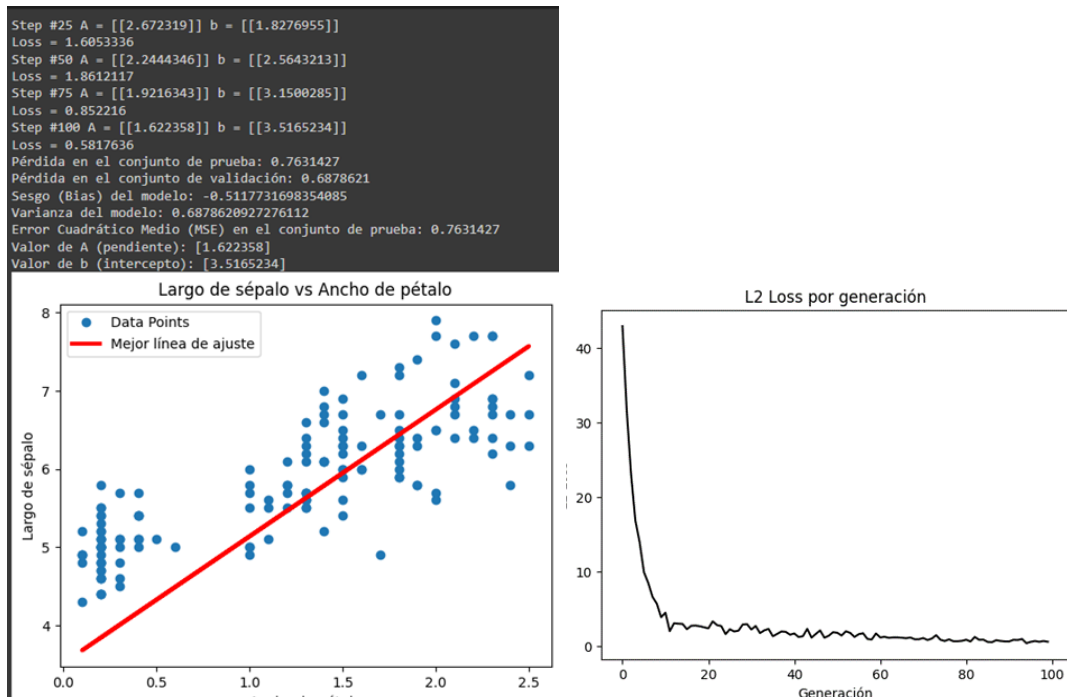


Imagen 1(LR = 0.03, epochs= 100)

Como se puede observar, el modelo tiene fallar en los datos más próximos origen, como mediciones este tenemos MSE el cual en este caso particular tiene un valor de 0.7631 y un sesgo de -0.511... lo que indica que el modelo tiende a subestimar valores objetivos siendo un mal modelo para este caso, con respecto al aprendizaje, la segunda gráfica muestra el desempeño del modelo y tal y como se puede observar, dicha gráfica tiende a 0,

Como parte de las mejoras para el modelo, modifiqué el valor del “Learning rate” para demostrar el valor que toma este elemento en el código:

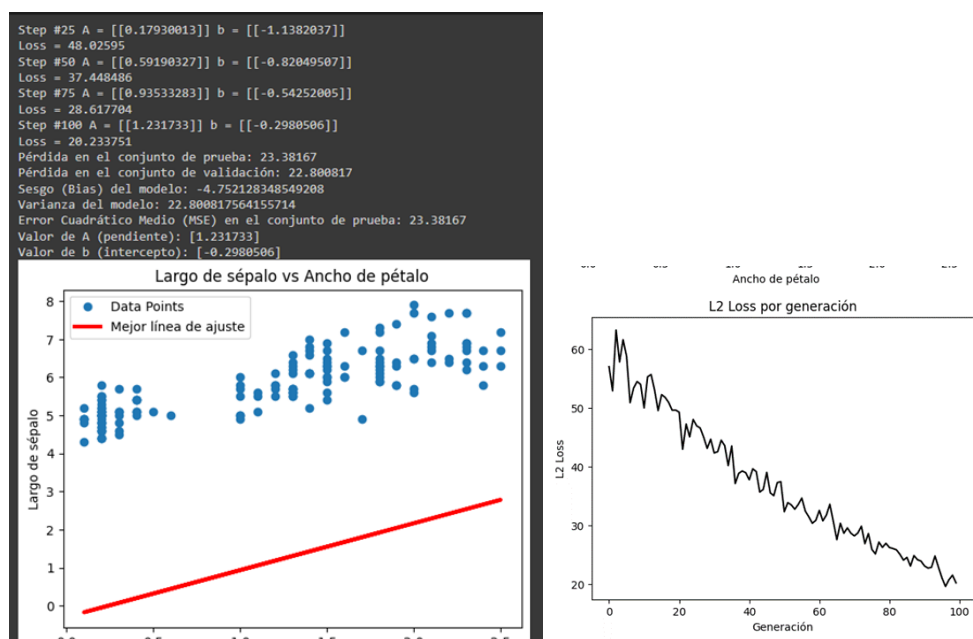


Imagen 2(LR = 0.001, epochs= 100)

Como se mencionó, se modificaron los valores del “Learning rate” y dado es, es evidente que la cantidad de epochs no le es suficiente para lograr un resultado mínimamente viable y tal y como se ve, el modelo está muy por debajo de los valores reales, por mencionar una métrica, el valor de MSE se encuentra en el rango de las decenas siendo de 22.8 con un sesgo de -4.25 por lo que para nada es útil, lo que nos da un claro ejemplo de “Underfitting” ya que el modelo obtenido (línea roja) no se acerca ni por un poco a los valores reales.

A continuación, se presenta la tercera prueba donde también se modificó el valor del “learning rate” por uno más alto siendo este de 0.05.

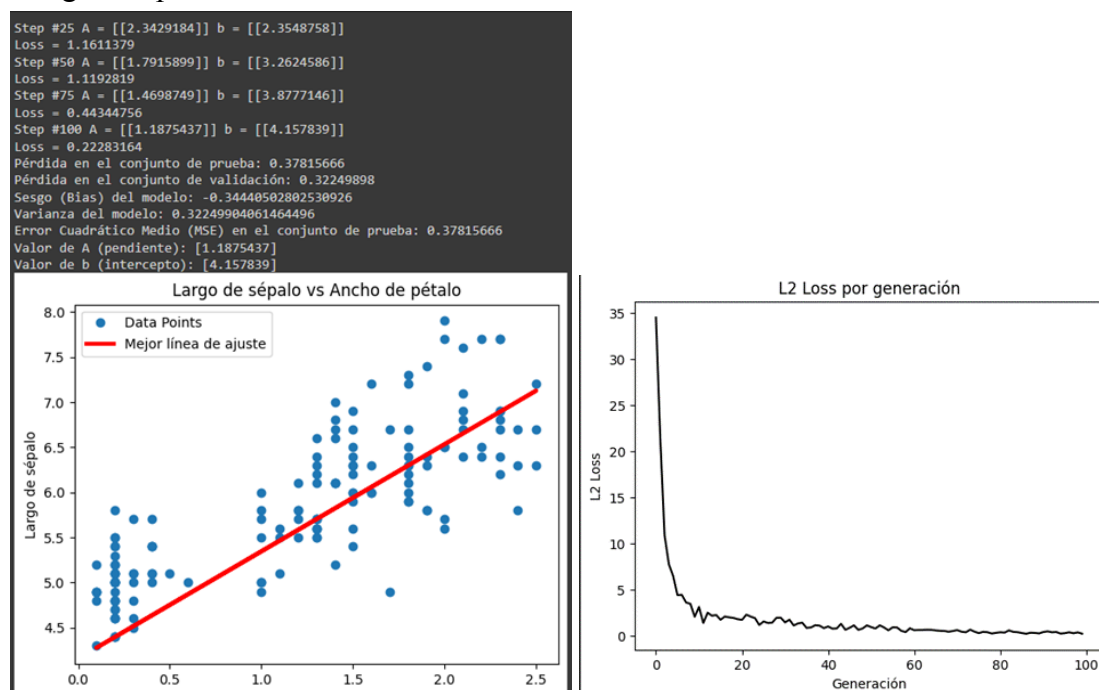


Imagen 3(LR = 0.05, epochs= 100)

Como se puede apreciar, la modificación en el valor nos mejores resultados de las dos pruebas previas, en primer lugar el valor del sesgo disminuye sin embargo aún es negativo, en el caso del MSE tenemos una baja drástica con respecto a los valores anteriores, en este caso es de 0.3781, con respecto al valor de la varianza en esta prueba resulta ser de las más altas con 0.032 lo que indica que la predicciones están dispersas alrededor de los datos reales. Para finalizar, con respecto al aprendizaje, podemos apreciar que el los valores de pérdida se precipitan de forma mas rapida ya que pasados los 40 epochs ya se tiene n valores cercanos al 0.

## Diagnóstico inicial

Dados los resultados anteriores, es claro que los modelos presentan un comportamiento sumamente diferente al alterar los valores de aprendizaje o “Hiperparametros” sin embargo, el análisis realizado carece de un valor número muy importante como lo es el R2 mismo que va en una escala de 0 a 1 el cual es de gran utilidad para conocer si el modelo de regresión se

ajusta, cabe mencionar que los resultados anteriores se basaron en los datos tal y como viene en el dataset, por lo que además lo mencionado anteriormente también es necesario aplicar una normalización de datos para trabajar con un estándar en ellos.

Aunado a lo anterior, también es necesario agregar una técnica que regularización a nuestro modelo para evitar el “Overfit”, para este caso se implementó el uso de Regularización Ridge (L2) la cual agrega una penalización basada en el cuadrado de los coeficientes de los parámetros del modelo.

Como último punto a mencionar, es que el código usado para las implementaciones posteriores a este punto del documento ya implementa cada mejora previa además de condensar todo en una función que permite manipular los hiperparámetros así como el tamaño de test, train y val.

Con el diagnóstico inicial y los cambios previamente mencionados, procedemos a realizar las mismas pruebas bajo las nuevas condiciones:

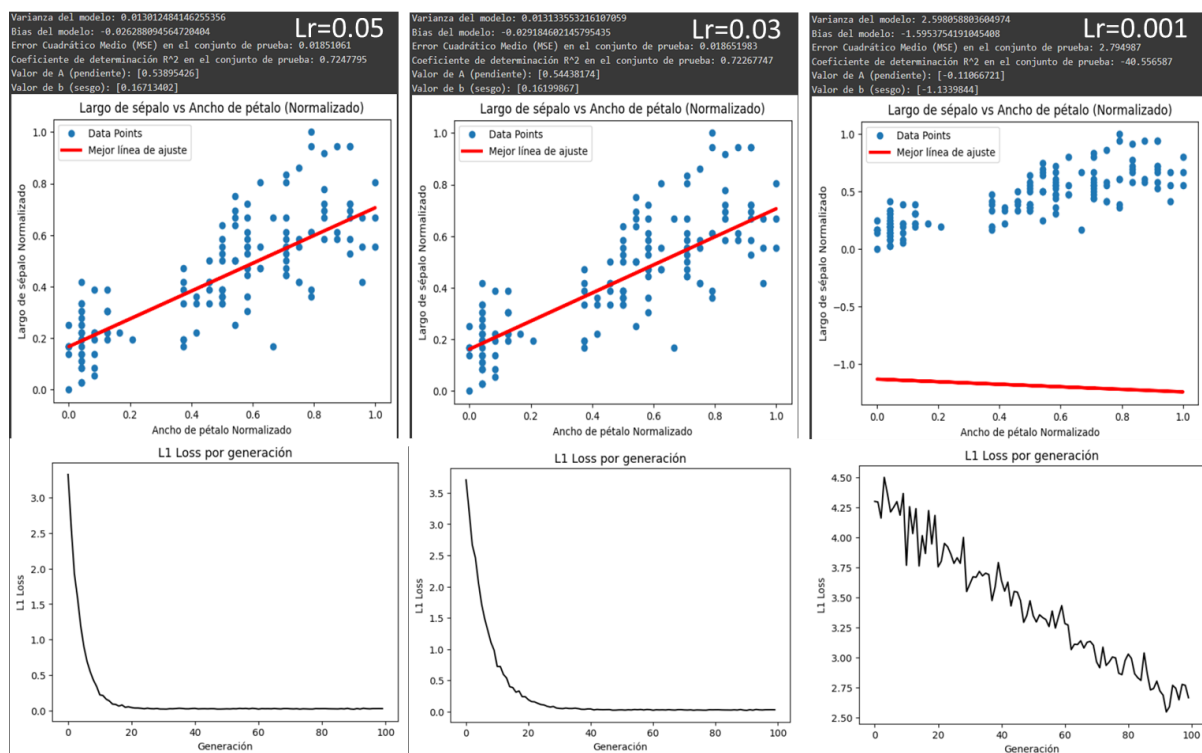


Imagen 4(epochs= 100,lambda\_reg = 0.01)

Varianza del modelo: 2.598058803604974 Bias del modelo: -1.5953754191045408 Error Cuadrático Medio (MSE) en el conjunto de prueba: 2.794987 Coeficiente de determinación $R^2$ en el conjunto de prueba: -40.556587 Valor de A (pendiente): [-0.11066721] Valor de b (sesgo): [-1.1339844]	<b>Lr=0.001</b>
Varianza del modelo: 0.013133553216107059 Bias del modelo: -0.029184602145795435 Error Cuadrático Medio (MSE) en el conjunto de prueba: 0.018651983 Coeficiente de determinación $R^2$ en el conjunto de prueba: 0.72267747 Valor de A (pendiente): [0.54438174] Valor de b (sesgo): [0.16199867]	<b>Lr=0.03</b>
Varianza del modelo: 0.013012484146255356 Bias del modelo: -0.026288094564720404 Error Cuadrático Medio (MSE) en el conjunto de prueba: 0.01851061 Coeficiente de determinación $R^2$ en el conjunto de prueba: 0.7247795 Valor de A (pendiente): [0.53895426] Valor de b (sesgo): [0.16713402]	<b>Lr=0.05</b>

Imagen 5 Valores reportados

Con respecto a los valores reportados en este set de pruebas, es claro que un learning rate más bajo(0.001) trae consigo consecuencias ligadas directamente con cantidad de epochs establecidos en el modelo, esto provoca que su varianza sea sumamente alta, hablando del bias este también es grande y se encuentre en el rango negativo indicando que los valores están consistentemente por debajo de los valores reales, eso junto el extremadamente bajo valor de R-cuadrada indican un obvio caso de “underfitting”.

Para el caso de las dos pruebas restantes, debo mencionar que son muy similares entres si, los valores reportados por diferentes por centésimas, aun así, los valores reportados con respecto a su varianza con en promedio 0.013 lo que indica que las predicciones del modelo son coherentes y tienden a agruparse cerca de un valor central el cual es un buen indicador del desempeño del modelo, en ambos casos el bias se encuentra en el rango negativo sin embargo al ser este muy bajo solo nos dice que se subestimamos los datos estimados a los reales.

Para finalizar, los valores de MSE y R-cuadrada son de 0.018 y 0.72 en promedio por lo que podemos decir con información matemática que el modelo es confiable para la evaluación de nuestra situación dado el contexto, cabe destacar que los valores reportados son sumamente similares a excepción del bias donde la prueba realizada con 0.05 de learning rate es 0.002 unidades menor.

Con todo lo anterior, decidí centrar otro tipo de pruebas para ver las capacidades del modelo bajo otro tipo de circunstancias por lo que continuaré usando el learning rate que mejor se desempeñó en todas las métricas calculadas (0.05). Con lo anterior claro, las siguientes pruebas modificarán el tamaño que tiene el test, train y val de una distribución de 60,20,20 a una de 80,10,10:

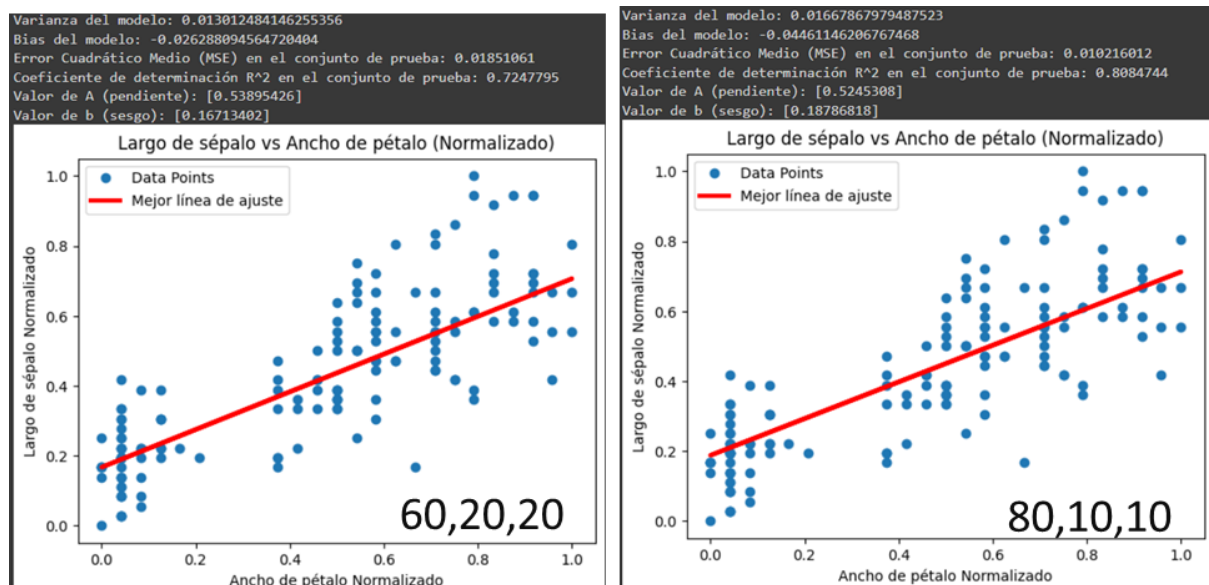


Imagen 6 Valores reportados de la modificación del tamaños

Con respecto a estos resultados puedo decir que si hay cambios significativos a comparación de las pruebas anteriores, comparando la varianza de ambos modelos es baja pero en la segunda prueba vemos un aumento de caso 0.003 unidades comúnmente una varianza baja es deseable pero analicemos los demás resultados. En el caso del bias podemos ver una aumentado de caso el doble en la segunda prueba indicando que el rango se subestimación del modelo es considerablemente grande con respecto al 60,20,20. En el caso de los últimos dos indicadores, podemos decir que la distribución 80,10,10 es superior con un valor de 0.08 unidades en el MSE y 0.008 unidades menos en el valor de R-cuadrada.

Como análisis general y en base a los valores obtenidos, sería muy fácil considerar que la distribución (80,10,10) es superior dado el valor de MSE y R-cuadrada sin embargo si analizamos mejor los valores de varianza y bias, podemos ver que la distribución 60,20,20 es la que menos subestima los valores reales, para ver mejor esto, tomemos los valores de A y b de ambas pruebas, si los comparamos numéricamente y gráficamente podemos ver que la pendiente de la primera prueba es mayor lo que nos dice que a la larga la estimación de los datos será mejor, sin embargo aquí es donde la normalización juega un rol definitorio ya que en este contexto, los datos han sido normalizados en una escala de 0 a 1 por lo que no nos interesa conocer valores fuera del rango previamente mencionado lo que le da la superioridad a la segunda distribución.

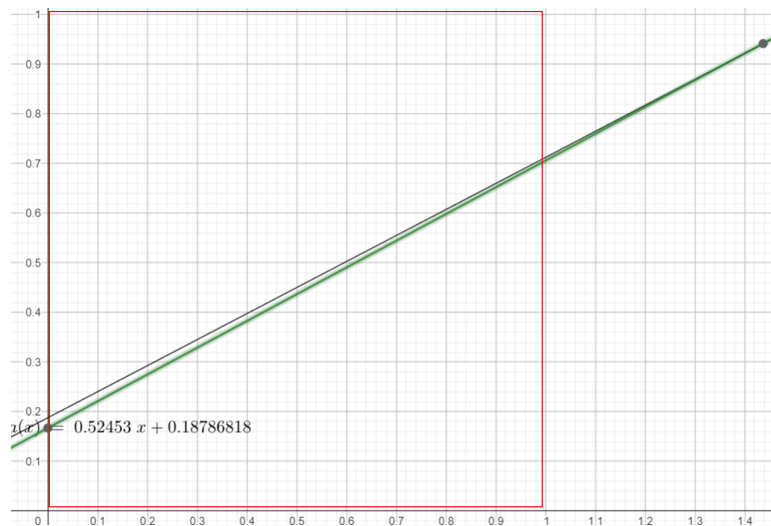


Imagen 7 Valores graficados

Para finalizar, las últimas dos pruebas simplemente se colocaron el valor de  $\lambda_{reg}$  usando los 100 epochs, 0.05 de aprendizaje y una distribución de 80,10,10.

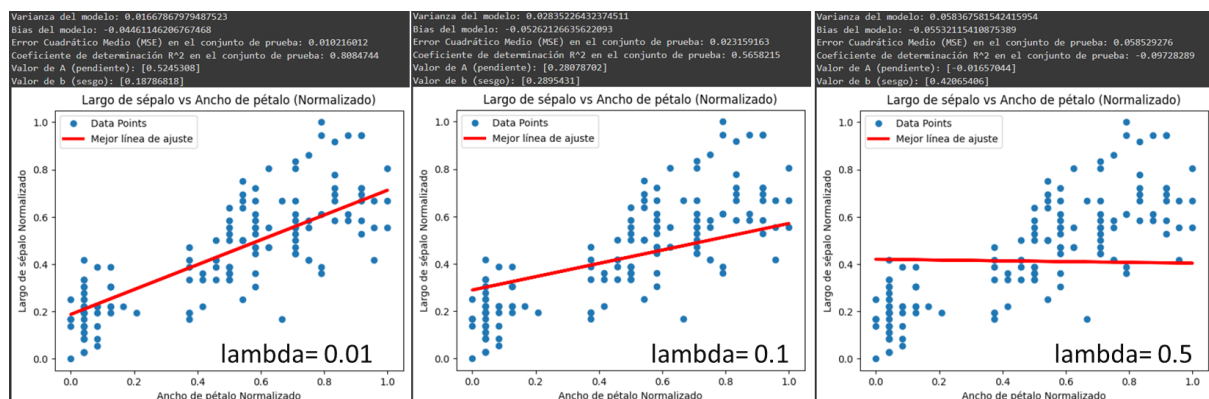


Imagen 8 Pruebas modificando el valor de lambda

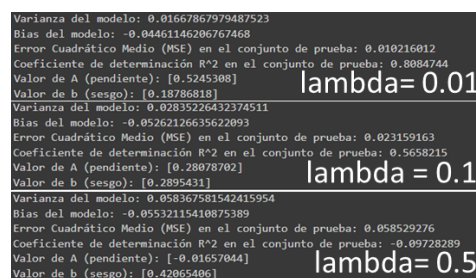


Imagen 9 Valores reportados

Con respecto a esta última prueba, podemos decir que las limitantes que pone el valor de  $\lambda$  al aprendizaje tiene considerables repercusiones en el modelo, en de forma experimental, podemos decir que que un valor bajo muy bajo no tiene muchas afectaciones ya que es con el que se ha trabajado la segunda mitad de este reporte, sin embargo al probar

valores ascendentes, podemos ver que el valor está castigando demasiado al aprendizaje lo que le impide reconocer patrones dentro de los valores, considero que en el primer test, se tiene un caso de donde el modelo ya puede considerar un “fitt” ya que los valores tales como el MSE y R-cuadrada, nos muestran que su desempeño es considerablemente bueno con respecto a las otras pruebas realizadas en este ejercicio, en el caso de las demás pruebas podemos decir que no reflejan para nada el comportamiento y evolución de los datos por lo que estas pruebas se realizaron con un enfoque demostrativo de los efectos de L2 en el modelo y como es su uso en la prevención del "overfitting".