



Inteligencia artificial avanzada para la ciencia de datos I

Momento de Retroalimentación: Módulo 2 Uso de framework o biblioteca de aprendizaje máquina para la implementación de una solución. (Portafolio Implementación)

Sobre el código

Modelo realizado

Para esta entrega se realizó un modelo de regresión lineal múltiple que puede recibir una cantidad n de características de entrada y una de salida. Se utiliza un formato csv para cargar los datos. El usuario puede elegir cuál es su característica de salida, y seleccionar cuales características quiere usar como entrada (o usar todas las restantes).

Hiperparámetros ajustables

En esta implementación no se tienen hiperparámetros ajustables.

Pruebas

Datasets de pruebas

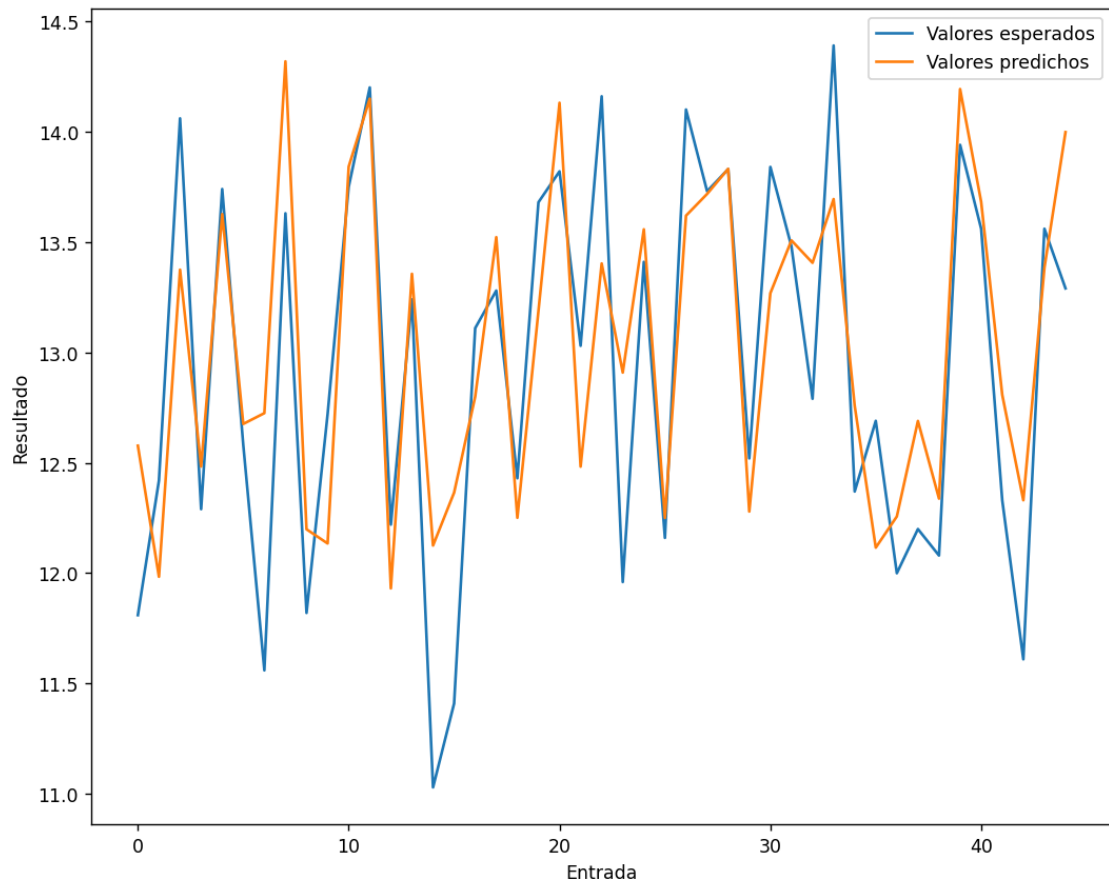
Para las pruebas se incluyó un dataset de prueba llamado wine.csv que son datos limpios.

Prueba con distintas columnas de entrada

El hecho de que nosotros podamos elegir las características de entrada nos permite obtener diferentes modelos con distinto nivel de accuracy o menor mse. Para estas pruebas vamos a tratar de predecir la columna *Alcohol*.

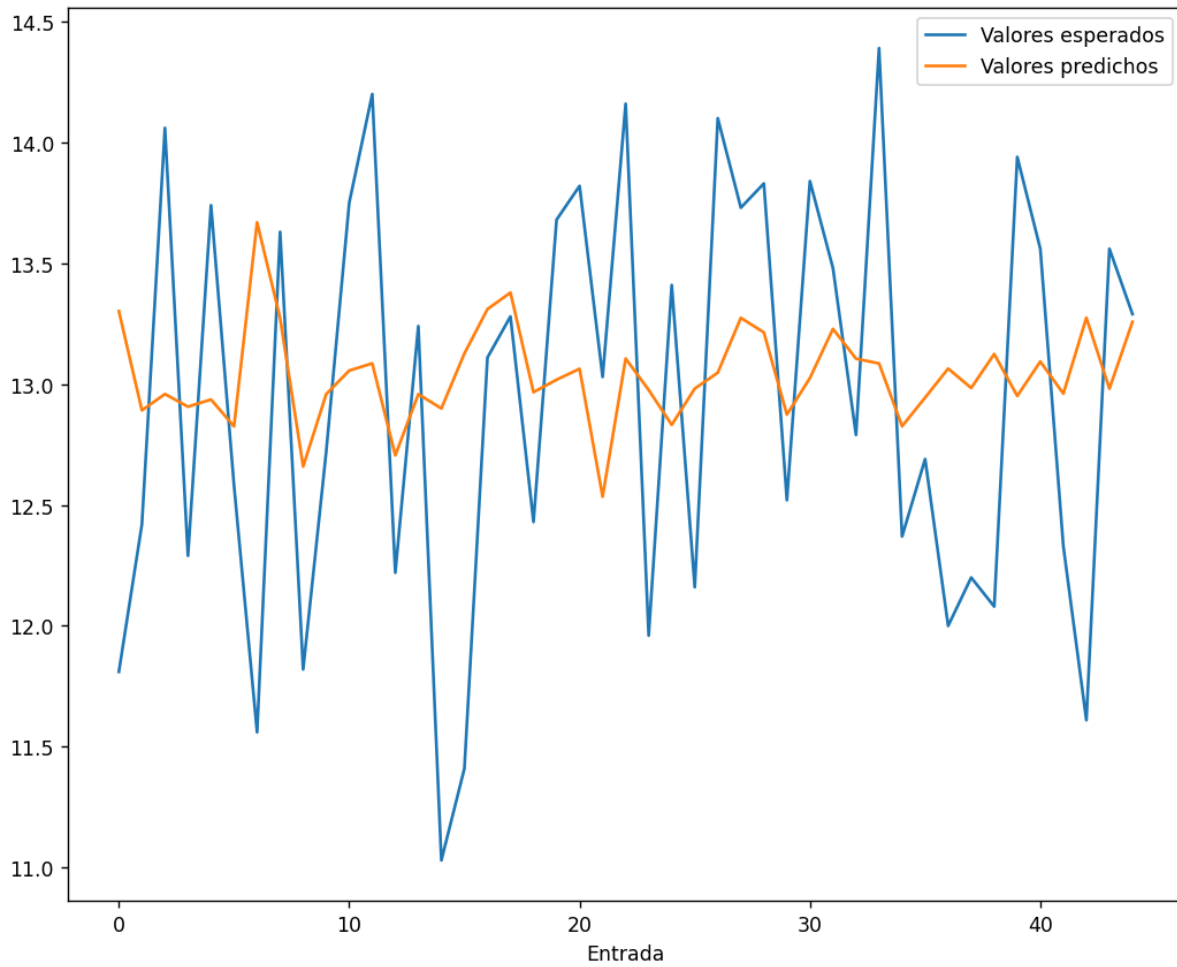
Prueba con todas las entradas

Se utilizó como característica de salida la columna *Alcohol* y se utilizaron las demás características de entrada.



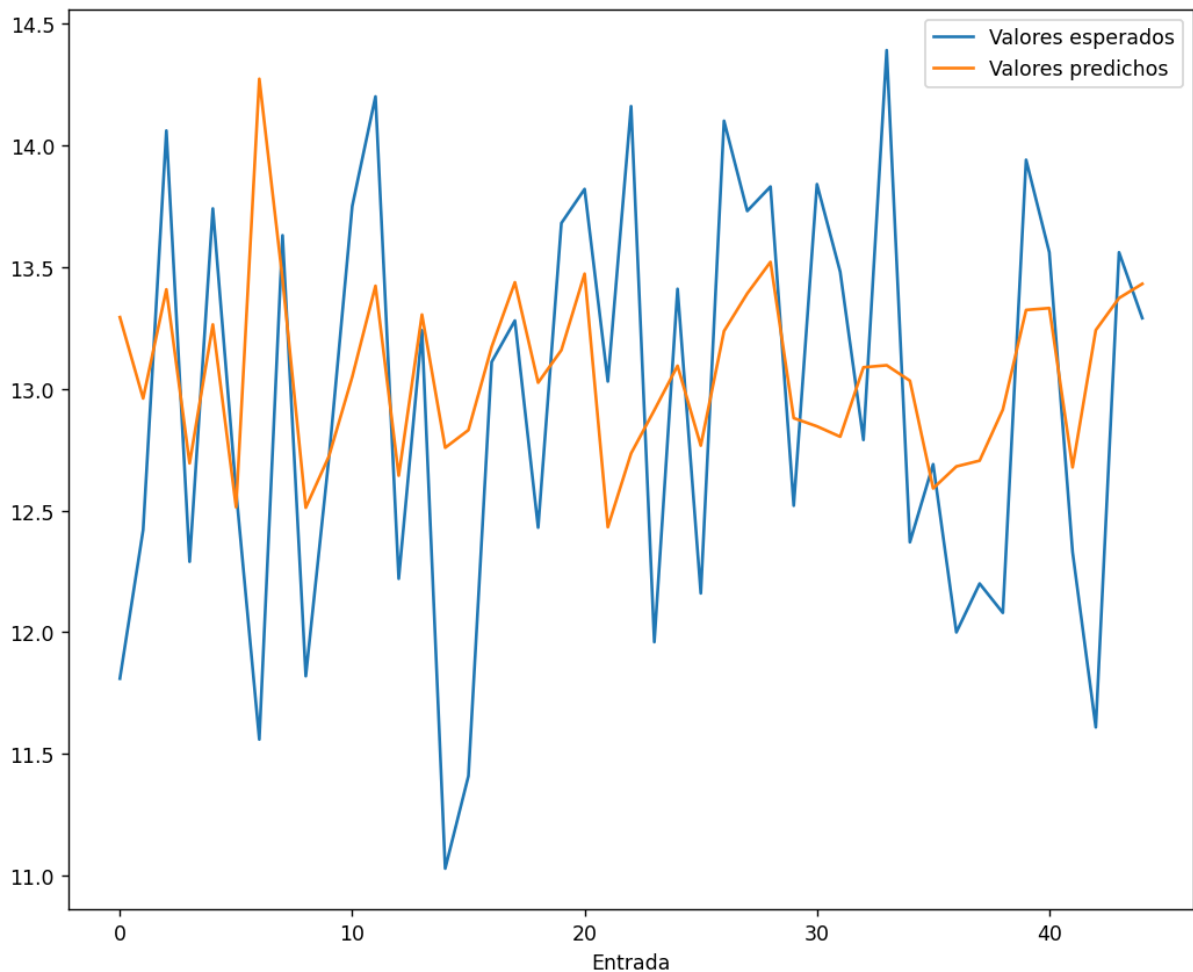
Lo cual nos dió un modelo con una R cuadrada ajustada de 0.531 y un MSE de 0.26364021976786634.

Prueba con 2 características de entrada



Esto nos dio un modelo con una R cuadrada ajustada de 0.056 y un MSE de 0.7882403474586703

Pruebas con 4 características de entrada



Esto nos da un modelo con una R cuadrada ajustada de 0.136 y un MSE de 0.6983882193388877

Análisis de accuracy y error

Información otorgada por el programa al finalizar

Al finalizar una corrida, se nos da información sobre el MSE de los datos con respecto a los resultados esperados para tener una idea de la precisión del modelo. También se nos proporciona la R cuadrada ajustada y valores como el p value de cada una de las características para poder saber cuáles características realmente son relevantes para el modelo.

Conclusiones importantes sobre el modelo

El modelo funciona muy bien sin hacer overfitting de los datos (hablando de los datos de prueba).

Algo muy notorio es que al modelo hay que introducirle datos que ya hayan sido previamente tratados, de lo contrario podría comenzar a realizar predicciones inexactas o simplemente no funcionar.