

## UD02\_01 Práctica Predicción desempeño académico

Sistemas de Aprendizaje Automático.

Álvaro Martínez Lineros

### Análisis del problema.

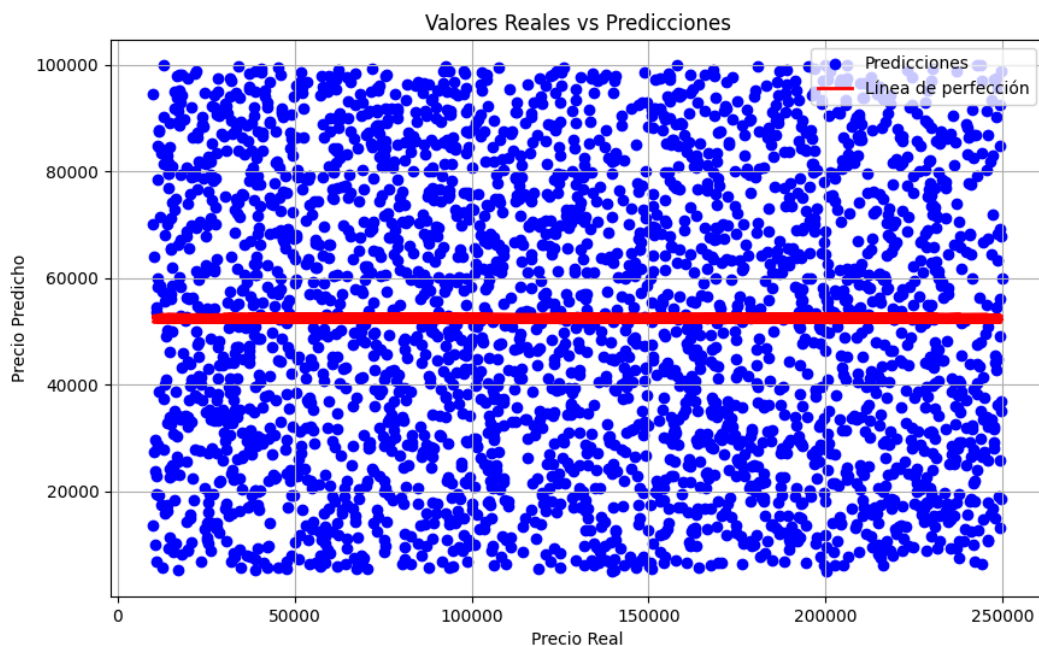
El objetivo de este estudio es desarrollar un modelo de clasificación que prediga el precio de unos vehículos teniendo datos como: motor, caballos, combustible, etc. La clasificación se realizará en tres categorías de rendimiento: "Bajo", "Medio" y "Alto".

### Estudio preliminar de los datos.

Al hacer un recuento y exploración inicial de los datos se ha visto que hay una gran variedad en los datos, donde no se ve una relación clara entre ningún factor y el precio.

### Regresión Lineal.

Se entrenó un modelo utilizando el algoritmo de Regresión Lineal con el dataset proporcionado y se puso a prueba con una parte de los mismos (80% datos de entrenamiento, 20% datos para prueba). Se obtuvieron los siguientes resultados:



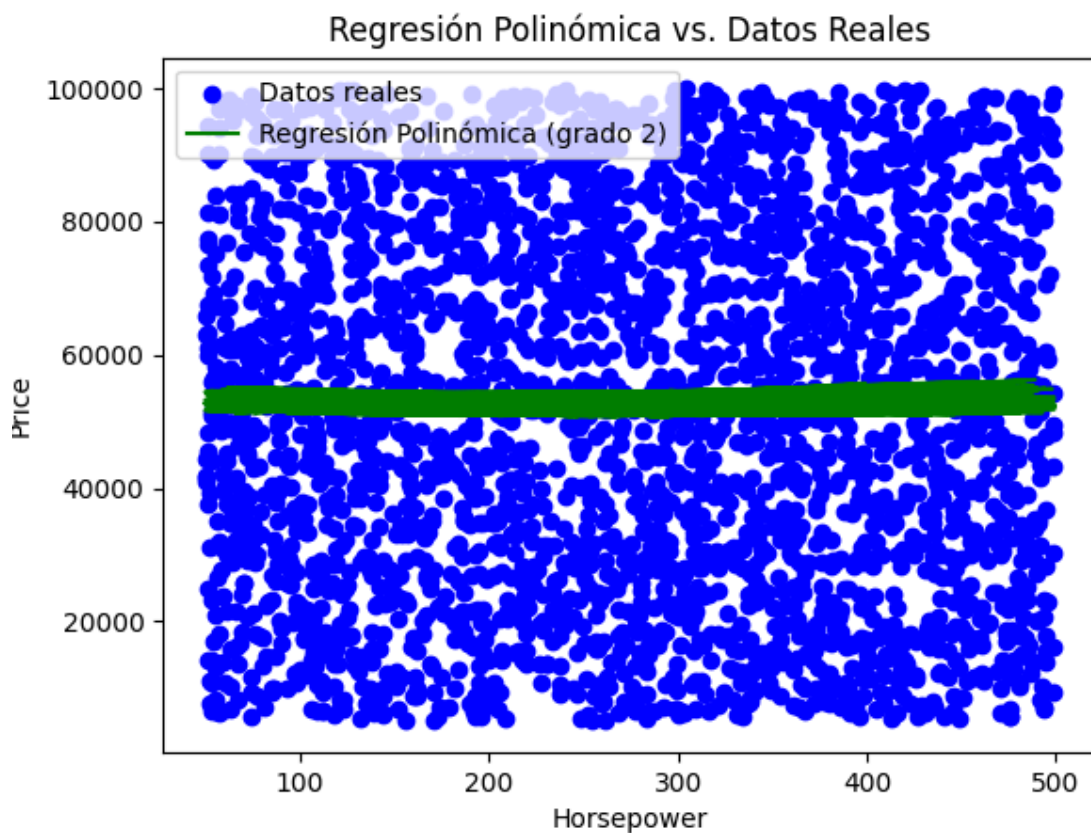
- Error absoluto medio(MAE): 23677.73
- Error cuadrático medio(MSE): 744728157.00
- Coeficiente de determinación ( $R^2$ ): 0.00

### Interpretación de los resultados.

El MAE obtenido es excesivamente alto, indica el error promedio entre los valores reales y los predichos. El MSE obtenido es también demasiado alto, por lo que el cuadrado de los errores está también muy alejado de los valores reales. Finalmente el  $R^2$  es 0, por lo que el modelo no es usable para la predicción de los precios.

### Regresión polinómica.

Se entrenó un modelo utilizando la regresión polinómica con el dataset proporcionado y se puso a prueba con una parte de los mismos (80% datos de entrenamiento, 20% datos para prueba). Se obtuvieron los siguientes resultados:



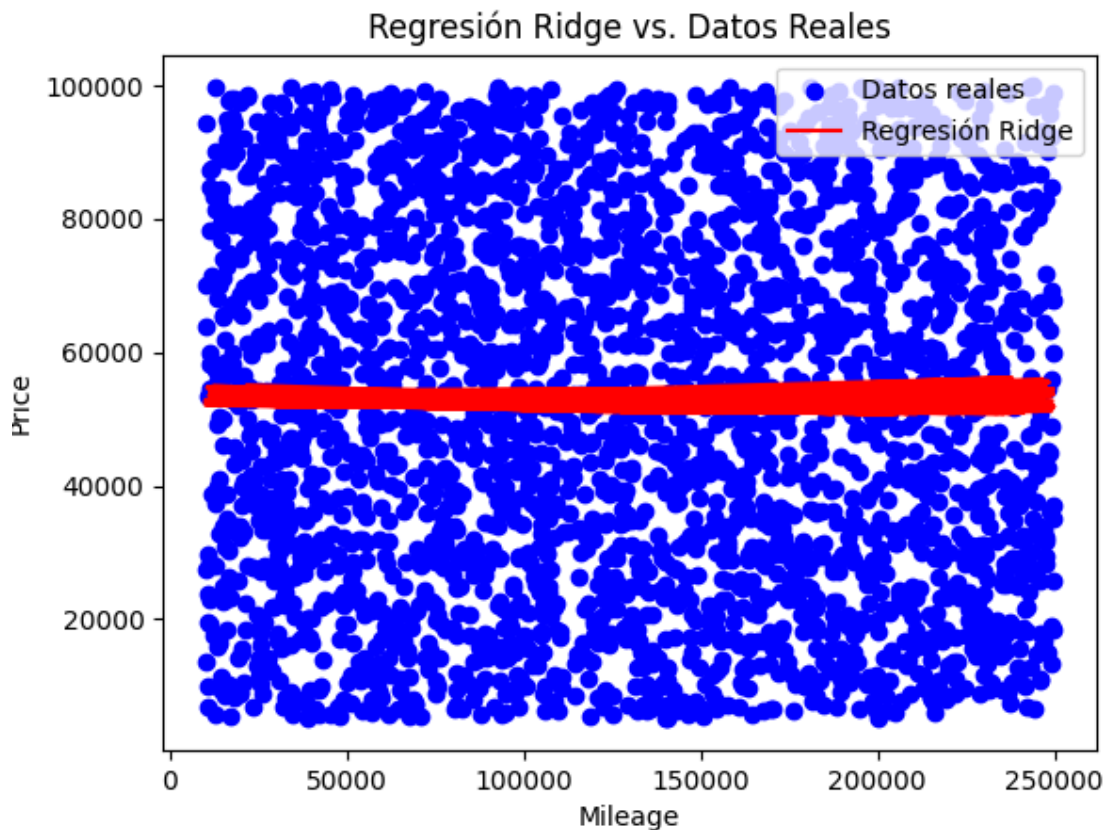
- Error absoluto medio(MAE): 23670.57
- Error cuadrático medio(MSE): 744562715.43
- Coeficiente de determinación ( $R^2$ ): 0.00

### Interpretación de los resultados.

El MAE obtenido es excesivamente alto, indica el error promedio entre los valores reales y los predichos. El MSE obtenido es también demasiado alto, por lo que el cuadrado de los errores está también muy alejado de los valores reales. Finalmente el  $R^2$  es 0, por lo que el modelo no es usable para la predicción de los precios.

## Regresión Ridge.

Se entrenó un modelo utilizando la regresión Ridge con el dataset proporcionado y se puso a prueba con una parte de los mismos (80% datos de entrenamiento, 20% datos para prueba). Se obtuvieron los siguientes resultados:



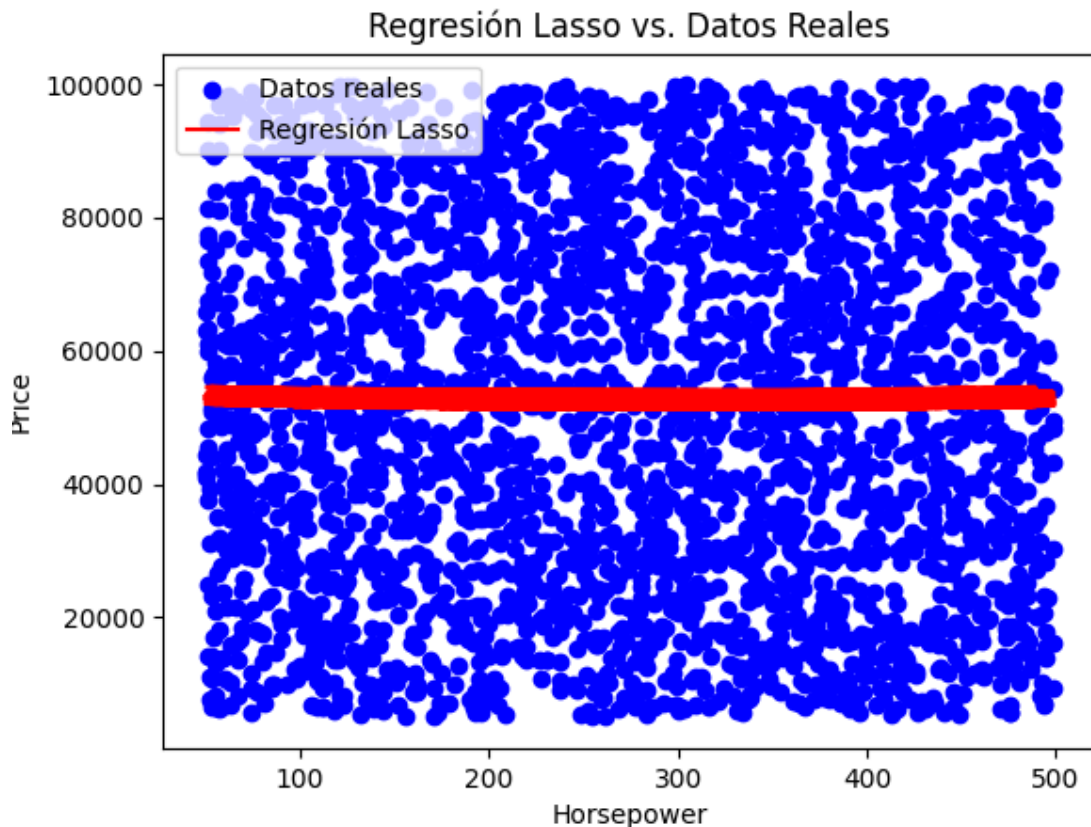
- Error absoluto medio(MAE): 23670.82
- Error cuadrático medio(MSE): 744585234.20
- Coeficiente de determinación ( $R^2$ ): 0.00

### Interpretación de los resultados.

El MAE obtenido es excesivamente alto, indica el error promedio entre los valores reales y los predichos. El MSE obtenido es también demasiado alto, por lo que el cuadrado de los errores está también muy alejado de los valores reales. Finalmente el  $R^2$  es 0, por lo que el modelo no es usable para la predicción de los precios.

## Regresión Lasso.

Se entrenó un modelo utilizando la regresión Ridge con el dataset proporcionado y se puso a prueba con una parte de los mismos (80% datos de entrenamiento, 20% datos para prueba). Se obtuvieron los siguientes resultados:



- Error absoluto medio(MAE): 23688.33
- Error cuadrático medio(MSE): 745187191.79
- Coeficiente de determinación ( $R^2$ ): 0.00

#### **Interpretación de los resultados.**

El MAE obtenido es excesivamente alto, indica el error promedio entre los valores reales y los predichos. El MSE obtenido es también demasiado alto, por lo que el cuadrado de los errores está también muy alejado de los valores reales. Finalmente el  $R^2$  es 0, por lo que el modelo no es usable para la predicción de los precios.

#### **Comparación de los resultados.**

Ninguno de los modelos entrenados es capaz de predecir el precio de los vehículos. Esto es porque los datos no presentan ninguna relación entre las variables, el precio no se ve variado de forma lineal o polinómica en ninguna de sus características.

Habría que hacer una revisión de los datos y comprobar que se han extraído correctamente. Eliminando valores extremos (outliers) y retirando cuartiles, tampoco se consigue que los modelos aprendan la predicción del precio.