COMPARACIÓN DE RESULTADOS DE MODELOS DE REGRESIÓN

Jairo Sánchez

2023-12-23

Introducción

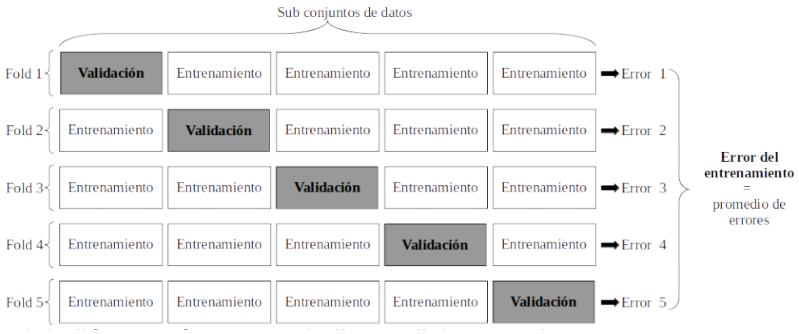
En los modelos de regresión que se han planteado se han contemplado únicamente modelos lineales; en este orden de ideas, se presenta los dos últimos modelos de este orden, los cuales corresponden a métodos de regularización, estos parten del modelo de regresión lineal al que se adiciona un factor de contracción. Las regresiones Ridge (L2) y Lasso (L1) son los métodos de regularización que se soportan en la minimización del RSS (Suma de residuos al cuadrado) al cual se le agrega un factor de penalización (o factor de contracción). El soporte matemático del método de regularización no es objeto del presente documento, sin embargo, una fuente recomendada para consultar o conocer el soporte teórico se encuentra en el la sección 6.2 del libro An Introduction to Statistical Learning Segunda edición, de los autores *James, G., Witten, D., Hastie, T., & Tibshirani.*

El desarrollo de esta parte del documento se soporta también en la sección 6.5.2 del libro ya mencionado; en este se recomienda el uso del método "validación cruzada tipo k-fold" con el fin de encontrar el factor de penalización que presente el mínimo de error en el modelo. Adicionalmente, se destaca que se hace uso de la librería "glmnet".

5.1. Validación cruzada tipo k-fold

R.Chollet, F., & Allaire, J. J en su libro Deep Learning with R mencionan que cuando un conjunto de datos no tiene una cantidad considerable de observaciones es recomendable hacer uso de la validación cruzada. No es muy claro (por lo menos en mi caso, no he encontrado una "formula" que me permita conocer el dato) cuál es la cantidad de observaciones óptima para concluir que son suficientes para trabajar únicamente con los subconjuntos Train y Test. La literatura si menciona que se debe procurar un "equilibrio" entre la cantidad de dimensiones (variables) y la cantidad de observaciones. Si se plantea que tenemos alrededor de 39 observaciones por cada variable (incluida la dependiente) se podría concluir que son muy pocas ya que en el libro en mención, en la sección 4.3.4, se dice que este conjunto de datos (Boston) posee muy pocas observaciones, y que plantear un subconjunto de validación (Test) con 100 observaciones (en promedio) es insuficiente, pues podría suceder que este subconjunto no represente adecuadamente el nivel de varianza que incluye el conjunto de entrenamiento (Train) y viceversa. En aquellos modelos de regresión en lo que se requiere encontrar el el optimo de uno o varios parámetros de tal manera que se minimice el error total del modelo,y que a su vez se tengan pocas observaciones, debería hacerse uso de la validación cruzada.

Existen varios métodos de validación cruzada, no obstante, para este planteamiento se usará la validación cruzada tipo k-fold, este consiste en dividir el conjunto de entrenamiento en k "pliegues" a fin de entrenar el modelo k veces, cada k entrenamiento entregará un valor de error (según la métrica de evaluación seleccionada), al final se hará un promedio de los k errores y este será el error de entrenamiento. En cada k entrenamiento se usa uno de los k pliegues para validar el entrenamiento que se hace con los k-1 pliegues restantes. Para un valor K = 5 lo anterior se observaría así:

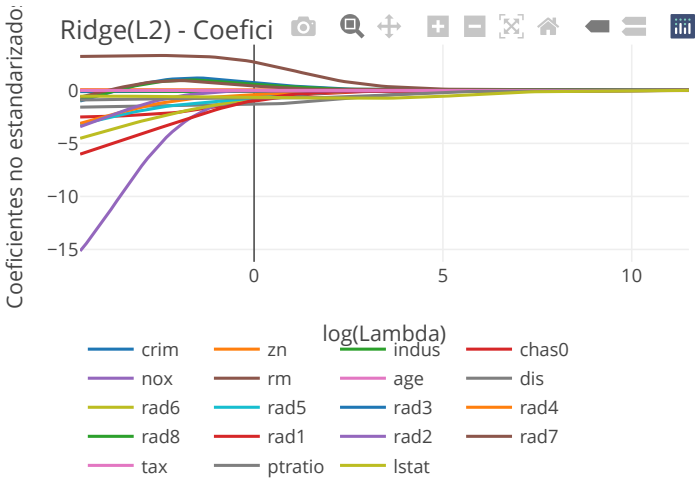


5.2. Regresión Ridge (L2)

El método de regularización ridge (L2) incluye, como factor de penalización en el RSS, la sumatoria de los coeficientes (de cada variable) al cuadrado y multiplicado por un valor lambda. En este método no se presenta "eliminación" de variables, para aquellas variables no influyentes el resultado de la aplicación de la penalización reduce el valor del coeficiente pero no se excluye, esto es, el resultado final va a incluir todas las variables independientes.

$$\text{RSS} + \lambda \sum_{j=1}^{p} \beta_j^2$$

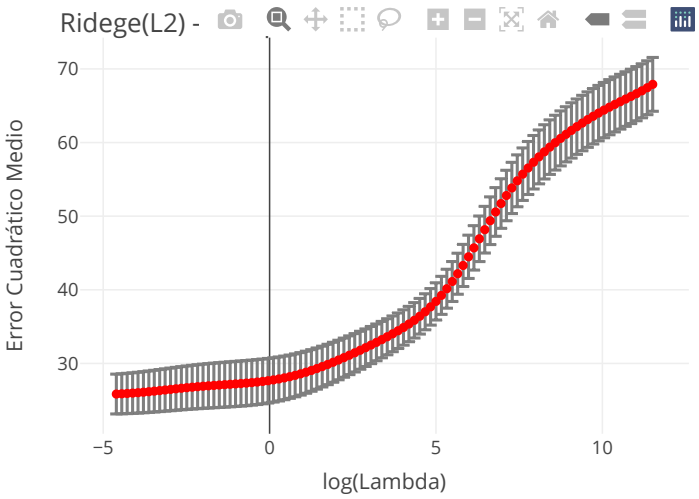
Para 100 valores lambda que están entre 0,001 y 100.000, los coeficientes de cada modelo (100) modelos se comportan como se observa en la siguiente gráfica:



Cuando el valor de lambda es pequeño la dispersión de los coeficientes se hace más evidente para las variables *nox* y *rm*, no obstante la primera desciende rápidamente a medida que el lambda se acerca a 1, el caso de *rm* tiene un descenso importante para un lambda mayor a 1. Para el resto de variables los valores de

los coeficientes se mantienen menos dispersos. Los valores de los coeficientes (o el lambda adecuado) que minimiza el error se obtiene mediante el cálculo del error medio cuadrado haciendo uso de la validación cruzada. Algunas de las variables que se desprenden de *rad* presentan un leve aumento en el valor de sus coeficientes para valores de lambda entre 0.02 y 0.1 y decaen hasta encontrar coeficientes cercanos a cero

Con un número de fols de 5 en la validación cruzada y los mismo valores lambda ya, el error cuadrado medio se observa así:

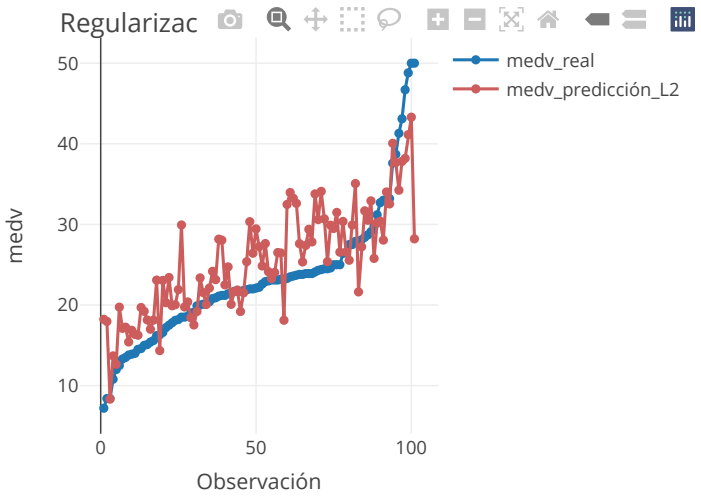


El menor MSE encontrado alcanza un valor de 25.83 con una desviación estándar de 4.08, el cual se encuentra para el valor lambda más bajo de los seleccionados, esto es, 0.01.

El modelo de regresión lineal múltiple con la regularización L2 con lambda de 0.01 se presenta así:

```
##   (Intercept)          crim            zn         indus         chas0
##  4.796339e+01 -1.139596e-01  6.669817e-02 -1.529567e-02 -2.508860e+00
##           nox            rm           age           dis          rad6
## -1.514274e+01  3.198992e+00  9.196584e-04 -1.582162e+00 -4.519528e+00
##          rad5          rad3          rad4          rad8          rad1
## -3.197726e+00 -1.001285e+00 -3.112311e+00 -9.003455e-01 -6.015038e+00
##          rad2          rad7           tax       ptratio         lstat
## -3.423904e+00 -6.363681e-01 -7.647806e-03 -9.061499e-01 -5.510993e-01
```

En este modelo, los coeficientes toman el valor más alto dentro de los 100 planteados (uno para cada valor lambda). La predicción de los valores para medv y su comparación con los valores reales se observaría de la siguiente manera:



En términos generales, los resultados de la predicción se observan por encima de los valores reales con una tendencia a quedarse por debajo9 cuando el valor medv real es de 38. Con lo observado gráficamente, el resultado de este modelo no será mejor que los analizados hasta el momento, en especial los modelos que

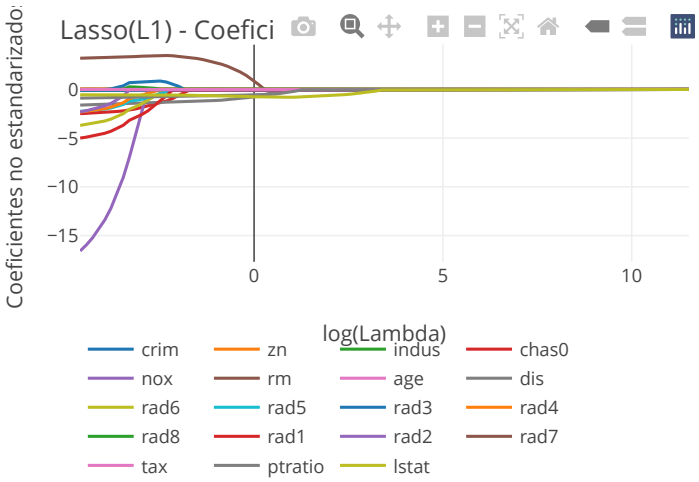
incluyen interacciones entre variables.

5.3. Regresión Lasso (L1)

Al igual que el modelo L2, Lasso (Least Absolute Shrinkage and Selection Operator) incluye un factor de penalización en el RSS: la sumatoria del valor absoluto de los coeficientes (de cada variable) multiplicado por un valor lambda. Dada la configuración de este factor de penalización, se da la posibilidad que exista "eliminación" de variables no influyentes.

$$\text{RSS} + \lambda \sum_{j=1}^{p} |\beta_j|$$

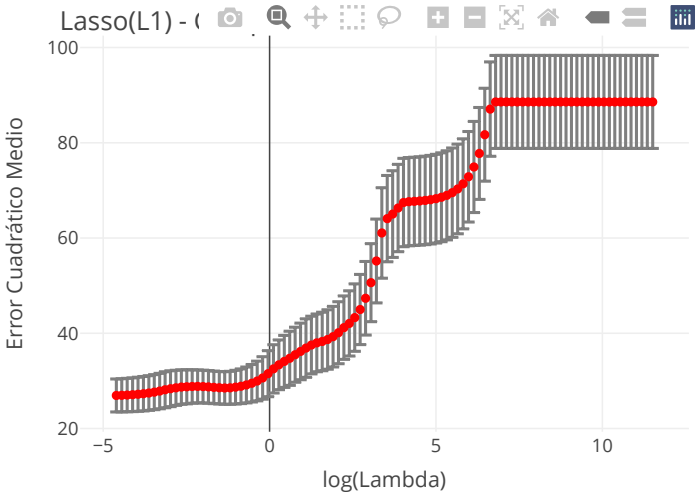
Se hace uso de los mismos cien valores lambda planteados en L2 por lo cual también se tendrán cien modelos lineales con un comportamiento en los coeficientes tal como lo presenta la siguiente gráfica:



De manera similar a la regresión ridge, para el valor más bajo de Lambda los valores de los coeficientes son altos para *nox* y *rm*, no obstante decrecen de manera rápida ubicándose muy cerca de cero. Para un valor lambda de 1 ya se observan un número considerable de variables independientes que tienden a desaparecer

del modelo de regresión lineal. Se destaca que en este caso las variables desglosadas *rad* pierden la tendencia a aumentar sus valores de coeficientes y luego a disminuir, a excepción de *rad3* entre valores lambda de 0.02 y 0.1 aproximadamente.

En cuanto al error cuadrado medio para cada una de las 100 opciones planteadas con L1, gráficamente se observan así:



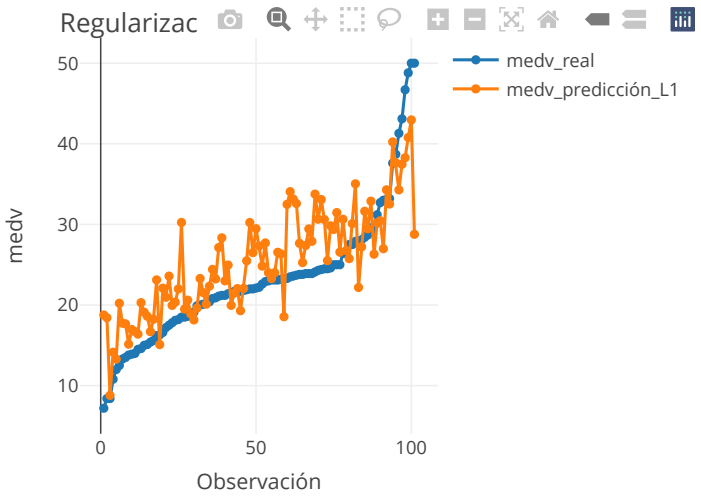
El comportamiento de los MSE es muy similar para el modelo L2 teniendo también como resultado el menor error cuando lambda es igual a 0.01, este es el menor valor de lambda dentro de los 100 valores entregados al modelo. el valor MSE es de 25.94 con una desviación estándar de 2.9. Si bien este error es ligeramente

alto respecto al observado con L2, el valor de la desviación se reduce en algo más de una unidad.

Con estos resultados se establece el modelo L1 con un valor lambda de 0.01

```
##   (Intercept)          crim            zn         indus         chas0
##  46.789356678  -0.108519863   0.066155690  -0.017603801  -2.488477475
##           nox            rm           age           dis          rad6
## -16.581946395   3.206550649   0.002073554  -1.605096329  -3.691830634
##          rad5          rad3          rad4          rad8          rad1
##  -2.250719475   0.000000000  -2.315010602   0.000000000  -4.991672592
##          rad2          rad7           tax       ptratio         lstat
##  -2.300139970   0.000000000  -0.005321277  -0.890659202  -0.548943024
```

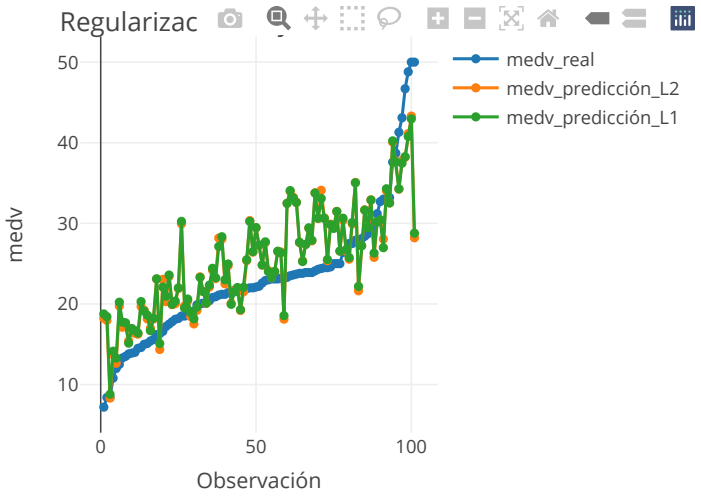
Tres de las variables incluidas en el modelo tienen como coeficiente cero, esto es, se excluyen del modelo final, *rad3, rad7* y *rad8*, las cuales se deprenden de la variable global *rad* . El modelo definitivo para L1 tiene un total de 16 variables.



El comportamiento de los datos de predicción de L1 tienen una tendencia similar al modelo L2, los datos de la predicción tienden a ubicarse por encima de los datos reales a excepción de los datos con valores de reales de medv superiores a 38 en los que la predicción genera un resultado evidentemente menor.

5.4. Comparación de modelos L1 y L2

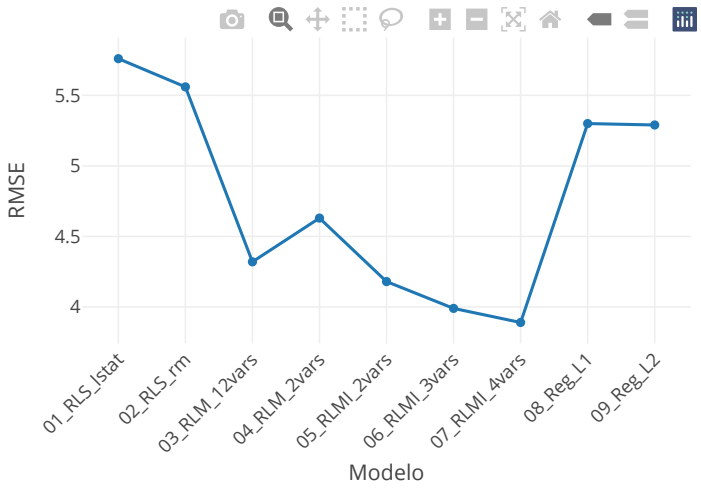
Los resultados de las predicciones de L1 y L2 se observan en la siguiente gráfica



Los modelos presentan resultados prácticamente igual, las diferencias en cada observación son mínimas entre L1 y L2, con esto se espera entonces que los resultados de RMSE sean muy cercanos.

5.4.1 Cálculo del RMSE Los resultados de RMSE para de ambos modelos tienen una diferencia de 0.01 lo que ratifica lo visto gráficamente: ambos modelos presentan resultados prácticamente iguales, la diferencia radica en que la complejidad del modelo L1 es menor al eliminar tres de las variables. Al compararlos con los resultados de RMSE con los de los demás modelos de regresión lineal desarrollados hasta el momento, estos son considerablemente altos, siendo solo un poco menor a los errores obtenidos en los modelos de regresión lineal simple, con base en lo anterior se puede concluir que para el conjunto de datos los modelos de regularización no son los más adecuados.

El RMSE del modelo L1 se calcula en 5.30 mientras que en el modelo L2 el valor es de 5.29, como ya se mencionó la diferencia es de 0.01, lo que representa un valor de 10 dólares , valor que es insignificante recordando que los datos de la variable *medv* se manejan en miles de dólares.


RMSE
5.5
5
4.5
4
01_RLS_lstat
02_RLS_rm
03_RLM_12vars
04_RLM_2vars
05_RLMI_2vars
06_RLMI_3vars
07_RLMI_4vars
08_Reg_L1
09_Reg_L2
Modelo