Tarea II
Tópicos Aplicados en Estadística

EYP3407

Nombre Diego Aravena
Alonso Campos

Problema 1

Asumiendo que $\boldsymbol{X}$ es ortogonal es decir, $\boldsymbol{X^T X} = \boldsymbol{I}$. obtenga el sesgo y la varianza de los estimadores de mínimos cuadrados, ridge y LASSO. Compárelos y comente los resultados.

Solución

Mínimos cuadrados ordinarios

Sabemos que la solución de mínimos cuadrados ordinarios es

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

Luego, podemos calcular tanto el sesgo y la varianza de los estimadores. Primero se calculará la esperanza del estimador para determinar si el estimador es insesgado.

$$\begin{aligned}
E(\hat{\boldsymbol{\beta}}) &= E\left((X^TX)^{-1}X^TY\right) \\
&= (X^TX)^{-1}X^TE(Y) \\
&= (X^TX)^{-1}X^TE(X\beta + e) && / \; Y = X\beta + e \\
&= (X^TX)^{-1}X^T[E(X\beta) + E(e)] \\
&= (X^TX)^{-1}X^T[X\beta + 0] && / \; e \sim N(0, \sigma^2 I) \\
&= (X^TX)^{-1}X^TX\beta \\
&= \beta
\end{aligned}$$

Por lo tanto, el estimador es insesgado, es decir,

$$Sesgo(\hat{\boldsymbol{\beta}}) = E(\hat{\beta}) - \beta = \beta - \beta = 0$$

Ahora calcularemos la varianza del estimador.

$$
\begin{aligned}
Var(\hat{\boldsymbol{\beta}}) &= Var\left((X^TX)^{-1}X^TY\right) \\
&= (X^TX)^{-1}X^TVar(Y)X(X^TX)^{-1} && \text{/ Forma cuadrática} \\
&= (X^TX)^{-1}X^TVar(X\beta+e)X(X^TX)^{-1} && / \; Y = X\beta + e \\
&= (X^TX)^{-1}X^T[Var(X\beta)+Var(e)]X(X^TX)^{-1} \\
&= (X^TX)^{-1}X^T[0+I_p\sigma^2]X(X^TX)^{-1} && / \; e \sim N(0,\sigma^2 I) \\
&= (X^TX)^{-1}X^TX(X^TX)^{-1}\sigma^2 \\
&= (X^TX)^{-1}\sigma^2 \\
&= I^{-1}\sigma^2 && \text{Como X es ortogonal, } X^TX = I \\
&= I\sigma^2
\end{aligned}
$$

Luego, la varianza para cada estimación corresponde a σ^2

Regresión Ridge

Sabemos que la solución de los estimadores Ridge son

$$\hat{\beta}^R_\lambda = (X^TX + \lambda I_p)^{-1}X^TY$$

Ahora determinaremos la esperanza del estimador

$$
\begin{aligned}
E(\hat{\beta}^R_\lambda) &= E((X^TX+\lambda I_p)^{-1}X^TY) \\
&= (X^TX+\lambda I_p)^{-1}X^TE(Y) \\
&= (X^TX+\lambda I_p)^{-1}X^T(E(X\beta)+E(e)) && Y = X\beta + e, \text{ con } e \sim N(0, I_p\sigma^2) \\
&= (X^TX+\lambda I_p)^{-1}X^T(X\beta+0) && Y = X\beta + e, \text{ con } e \sim N(0, I_p\sigma^2) \\
&= (X^TX+\lambda I_p)^{-1}X^TX\beta \\
&= (I+\lambda I_p)^{-1}I\beta && \text{X es ortogonal, } X^TX = I \\
&= (1+\lambda)^{-1}\boldsymbol{\beta}
\end{aligned}
$$

El cual corresponde a un estimador sesgado del parámetro, lo que se debe principalmente al parámetro de penalización λ.

$$
\begin{aligned}
Sesgo(\hat{\boldsymbol{\beta}}) &= E(\hat{\beta}) - \beta \\
&= (1+\lambda)^{-1}\boldsymbol{\beta} - \boldsymbol{\beta} \\
&= ((1+\lambda)^{-1}I_p - I_p)\boldsymbol{\beta} \\
&= ((1+\lambda)^{-1} - 1)I_p\boldsymbol{\beta} \\
&= \frac{-\lambda}{1+\lambda}I_p\boldsymbol{\beta}
\end{aligned}
$$

Posteriormente calculamos la varianza del estimador

$$
\begin{aligned}
Var(\hat{\beta}_\lambda^R) &= Var((X^TX + \lambda I_p)^{-1}X^TY) \\
&= Var((I_p + \lambda I_p)^{-1}X^TY) \\
&= Var((1+\lambda)^{-1}I_pX^TY) \\
&= (1+\lambda)^{-2}I_pVar(X^TY) \\
&= (1+\lambda)^{-2}I_pX^TX\sigma \\
&= (1+\lambda)^{-2}I_p\sigma \\
&= \left(\frac{\sigma}{1+\lambda}\right)^2 I_p
\end{aligned}
$$

De este modo, se concluye que el estimador Ridge tiene menor o igual varianza que el EMCO para cualquier valor de λ. A su vez, se destaca que al ser este penalizador igual a cero, se estaría llegando al mismo valor que el de mínimos cuadrados.

Regresión LASSO

El objetivo en este caso se trata de minimizar

$$
(Y - X\beta)^T(Y - X\beta) + \lambda\sum_{j=1}^{p}|\beta_j|
$$

Lo cual corresponde a minimizar la suma de cuadrados residuales sujeto a una penalización en l_1, por ende, esto lo vuelve un problema de optimización convexa. Luego, se busca encontrar $\hat{\beta}^L$ a través de la ecuación

$$
-2X^T(Y - X\beta) + \lambda\gamma = 0
$$

donde, $\gamma_j = signo(\beta_j)$ si $\beta_j \neq 0$ y $\gamma_j \in [-1, 1]$ si $\beta_j = 0$. En este caso como la matriz de diseño es ortogonal, vimos que el estimador de mínimos cuadrados corresponde a la matriz X^TY, por lo tanto, la ecuación para encontrar $\hat{\beta}_L$ equivale a determinar

$$-2\hat{\beta}^{mc} - 2\beta + \lambda\gamma = 0$$

Notemos que si $\beta_j = 0$, entonces

$$\begin{array}{ccc} -1 & \leq \gamma & \leq 1 \\ -\dfrac{\lambda}{2} & \leq \frac{\lambda\gamma}{2} & \leq \dfrac{\lambda}{2} \end{array}$$

Por ende, vemos que se cumple la condición $|\hat{\beta_j}^{mc}| \leq \frac{\lambda}{2}$. Luego, es posible deducir que si $|\hat{\beta_j}^{mc}| \leq \frac{\lambda}{2} \rightarrow \hat{\beta_j}^{L} = 0$.

Por otro lado, si $\beta_j \neq 0$, entonces

$$\begin{aligned} -2\hat{\beta_j}^{mc} - 2\beta_j + \lambda signo(\beta_j) &= 0 \\ \beta_j + \frac{\lambda signo(\beta_j)}{2} &= \hat{\beta_j}^{mc} \end{aligned}$$

Es decir,

$$\hat{\beta_j}^{L} = \begin{cases} \hat{\beta_j}^{mc} - \frac{\lambda}{2} & si \quad \hat{\beta_j}^{mc} > \frac{\lambda}{2} \\ \\ \hat{\beta_j}^{mc} + \frac{\lambda}{2} & si \quad \hat{\beta_j}^{mc} < -\frac{\lambda}{2} \end{cases}$$

Del mismo modo, podemos concluir que el estimador LASSO dependerá del valor que estime el EMCO.

Al calcular la esperanza del estimador LASSO se tiene lo siguiente

$$E(\hat{\beta_j}^{L}) = \begin{cases} \beta - \frac{\lambda}{2} & si \quad \hat{\beta_j}^{mc} > \frac{\lambda}{2} \\ \\ \beta + \frac{\lambda}{2} & si \quad \hat{\beta_j}^{mc} < -\frac{\lambda}{2} \\ \\ 0 & si \quad |\hat{\beta_j}^{mc}| \leq \frac{\lambda}{2} \end{cases}$$

Claramente se obtienen estimadores sesgados, en particular

$$Sesgo(\hat{\beta_j}^{L}) = \begin{cases} -\frac{\lambda}{2} & si \quad \hat{\beta_j}^{mc} > \frac{\lambda}{2} \\ \\ \frac{\lambda}{2} & si \quad \hat{\beta_j}^{mc} < -\frac{\lambda}{2} \\ \\ -\beta & si \quad |\hat{\beta_j}^{mc}| \leq \frac{\lambda}{2} \end{cases}$$

Por lo que podemos determinar el sesgo al cuadrado,

$$\left(Sesgo(\hat{\beta}_j^{\,L})\right)^2 = \begin{cases} \left(\frac{\lambda}{2}\right)^2 & si \quad |\hat{\beta}_j^{\,mc}| > \frac{\lambda}{2} \\ \\ \beta^2 & si \quad |\hat{\beta}_j^{\,mc}| \leq \frac{\lambda}{2} \end{cases}$$

Por último, la varianza de estos estimadores tienen la siguiente estructura

$$Var(\hat{\beta}_j^{\,L}) = \begin{cases} \sigma^2 & si \quad |\hat{\beta}_j^{\,mc}| > \frac{\lambda}{2} \\ \\ 0 & e.o.c \end{cases}$$

En síntesis, podemos notar que los métodos de regularización al considerar una penalización pierden la propiedad de ser insesgados. Sin embargo, éstos presentan menor variabilidad en las estimaciones, pues en el caso del método Ridge corresponde a la misma varianza del EMCO, pero ponderado por $1/(1+\lambda)^1$ lo que genera que disminuya la variabilidad de la estimación. En el caso de LASSO, en primer lugar notamos que los estimadores se denotan como una función por partes debido a su penalización mediante la norma 1. De igual manera, éstos presentan menor variabilidad, pues en un caso se tiene al misma incertidumbre que los estimadores EMCO, pero en los demás no hay variabilidad, ya que siempre será cero. En ambos casos, dependiendo de los valores que puedan tomar los parámetros, es posible que ambos estimadores puedan tener un menor error cuadrático medio que el estimador EMCO.

Además, cabe señalar que corresponde al caso en que la matriz de diseño sea ortogonal, recordemos que existen problemas de invertibilidad si existe altas correlaciones entre predictores, por lo que éstos estimadores adquieren aún mayor relevancia en dicho caso.

Problema 2

Descargue el conjunto de datos *hitters*. Esta base corresponde a los datos de una liga de baseball entre las temporadas de 1986 y 1987. Para una descripción de los datos puede ver el enlace `https://cran.r-project.org/web/packages/ISLR/ISLR.pdf`. La variable respuesta para este problema es **Salario**. Como la distribución del Salario es sesgada, se debe tomar la transformación $Y = \log(\text{Salario})$.

(a) ¿Cuáles son las características más importantes para predecir el salario de los jugadores?

   (i) Ajuste y visualice métodos de regularización vistos en clase (LASSO, Elastic-Net) incluyendo LASSO adaptativo.

   (ii) ¿Cuáles son los mejores predictores seleccionados por cada método? ¿Son diferentes?

(b) ¿Cuál es el método es mejor para predecir el salario de los jugadores? Para hacer la decisión considere un set de entrenamiento (60 %), validación (20 %) y test (20 %). Si los métodos considerados tienen parámetros que calibrar, entonces se debe ajustar el modelo con el set de datos de entrenamiento, se debe elegir el parámetro a calibrar minimizando el error de predicción en el set de validación y se debe reportar la predicción final con los datos de testeo. Se debe repetir este procedimiento 10 veces y reportar los resultados promedio.

(iii) Compare el MSE promedio obtenido en los set de datos de testeo considerando (a) mínimos cuadrados, (b) regresión Ridge, (c) LASSO, (d) Elastic-Net y (e) LASSO adaptativo.

(iv) Visualice los resultados obtenidos para comprar los modelos. Muestre los resultados solamente para el mejor parámetro de calibración elegido.

(v) ¿Qué métodos generan el mejor error de predicción? ¿Por qué estos métodos funcionan bien? ¿Los métodos eligen el mismo subconjunto de variables? Explique y amplíe sus respuestas.

Solución (a)

Análisis Exploratorio

El conjunto de datos `Hitters` contiene 322 registros y 21 variables con información de los jugadores tanto en la temporada 1986 como en su carrera completa. En la Tabla 1 se podrá observar la descripción de cada una de ellas.

En el estudio de la base de datos se hallaron 59 registros de jugadores que no contenían información en la variable respuesta, es decir, en la variable `Salary`. De este modo, se realiza un estudio de los valores faltantes.

Estudio valores faltantes

En el conjunto de datos no se logra obtener la información completa del salario de los jugadores, pues existen 59 registros faltantes en la base de datos, lo que equivale a un 18,32 % del total.

A continuación se estudió el comportamiento de algunas variables para determinar la causalidad de la aparición de datos faltantes, en particular la Figura 1 muestra aquellas variables asociadas al desempeño del jugador a lo largo de su carrera, donde se ve que no existen diferencias notorias en aquellos registros que contienen o no valores faltantes en `Salary`. Del mismo modo fueron considerados otros grupos de variables con ánimos de encontrar diferencias claras entre ambos grupos, no obstante, no fue posible obtener conclusiones claras sobre el hecho de que no se hayan obtenido aquellos registros. Por lo tanto, se decidió ignorar aquellas observaciones sin registros en la variable respuesta dado que bajo este contexto los problemas que surgen al disminuir el n son inferiores que en el contexto de mínimos cuadrados.

Variable	Descripción
`Jugador`	Nombre del jugador
`AtBat`	Número de veces que tomó el turno para batear en 1986
`Hits`	Número de *hits* en 1986
`HmRuns`	Número de *home runs* en 1986
`Runs`	Número de carreras en 1986
`RBI`	Número de carreras bateadas (carreras bateadas impulsadas) en 1986
`Walks`	Número de bases por bolas (*walks*) en 1986
`Years`	Número de años en las ligas mayores
`CAtBat`	Número de veces que tomó el turno para batear durante su carrera
`CHits`	Número de *hits* durante su carrera.
`CHmRuns`	Número de *home runs* durante su carrera.
`CRuns`	Número de carreras durante su carrera.
`CRBI`	Número de carreras bateadas (carreras bateadas impulsadas) durante su carrera.
`CWalks`	Número de bases por bolas (*walks*) durante su carrera.
`League`	Un factor con niveles A (American) y N (National) que indican la liga del jugador a finales de 1986.
`Division`	Un factor con niveles E (East) y W (West) que indican la división del jugador a finales de 1986.
`PutOuts`	Número de *put outs* en 1986
`Assists`	Número de asistencias en 1986
`Errores`	Número de errores en 1986
`Salary`	Salario anual de 1987 en el día de apertura, en miles de dólares
`NewLeague`	Un factor con niveles A y N que indican la liga del jugador a principios de 1987

Tabla 1: Descripción conjunto de datos `Hitters`

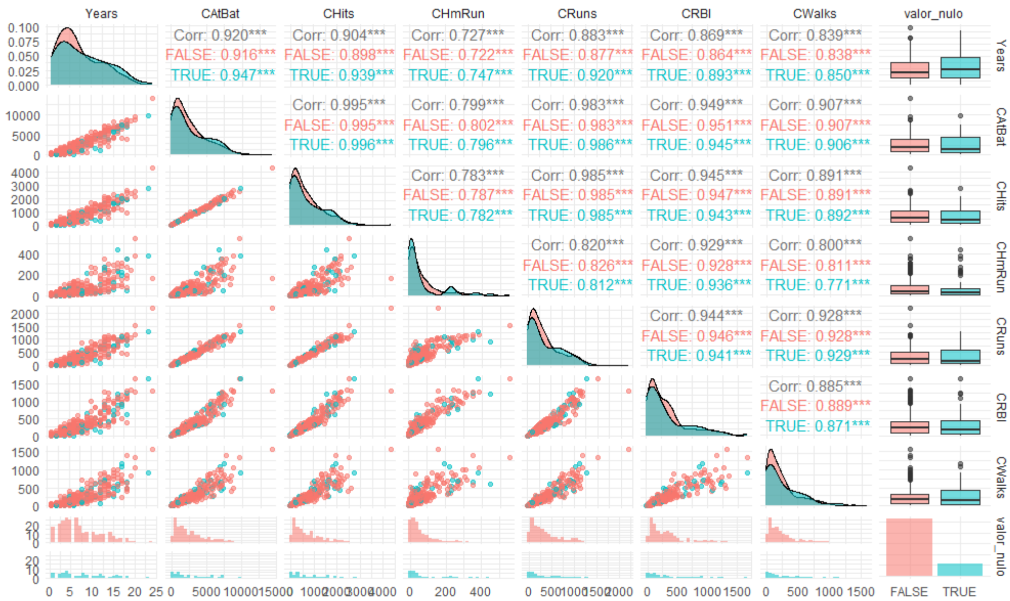


Figura 1: Comparación datos que contienen o no valor nulo en variable `Salary`

.

Estudio de correlaciones

Los métodos de regularización se utilizan principalmente cuando existe una gran cantidad de variables en modelo de regresión, ya que, en ese caso en la mayoría de las ocasiones éstos estarán correlacionados ente sí y por consiguiente afectará en la varianza estimaciones. Por este motivo, estudiaremos la correlación entre las variables.

En la Figura 2 se puede observar que existe un patrón en las correlaciones del conjunto de datos. En primer lugar, es posible notar que existe una alta correlación en las primeras variables, las cuales están ligadas con el rendimiento del jugador durante la temporada de 1986. Luego, se distingue otro grupo de variables altamente correlacionadas y que están asociadas a la carrera del jugador. De igual manera la variable `Salary` no mantiene altas correlaciones con las demás variables. Por otro lado, es posible advertir que el número de *puts outs* no está relacionado con las demás. Además, la cantidad de asistencias y errores del jugador en la temporada 1986 presentan una correlación alta.

Modelos de Regresión

Con el objetivo de determinar los predictores que más influyen en los salarios de los jugadores, se ajustarán cuatro modelos, éstos son las regresiones Ridge, LASSO, Elastic-Net y LASSO adaptativo.

En cada uno de éstos métodos, se consideran las variables estandarizadas en el caso de variables numéricas, mientras que para las variables categóricas se consideran variables

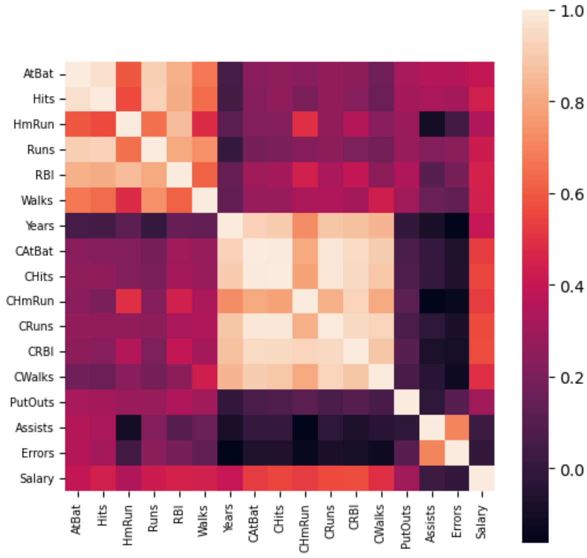


Figura 2: Correlaciones variables `Hitters`

dummies con celda de referencia. Además, para determinar el parámetro de penalización se utiliza validación cruzada utilizando 10 particiones en los datos.

Regresión Ridge

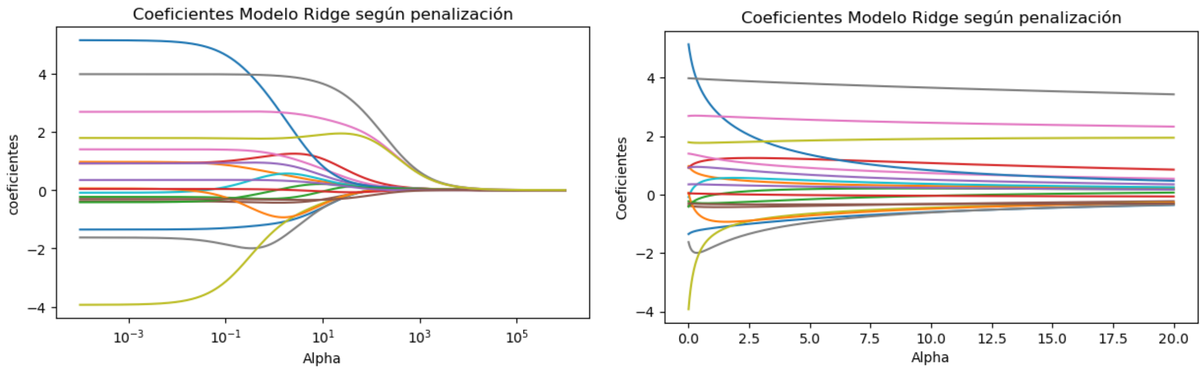
La Figura 3 muestra la traza de ridge de los coeficientes estimados por el modelo para distintos valores de α. Los coeficiente comienzan a estabilizarse para un α cercano a 10. Este α corresponde al penalizador que minimiza el Error Cuadrático Medio, en particular, en este caso su valor óptimo es 10.12.

Recordemos que en la regresión Ridge usualmente los coeficientes estimados no son cero. En la Figura 4 se concluye que las variables que adquieren mayor influencia sobre el salario de los jugadores son `Hits` y `years`, mientras las que resultan menos relevantes son `HmRun`, `RBI` y `CHmRun`.

Regresión LASSO

Al igual que en el caso anterior, se ajusta un modelo de regresión, esta vez con el método LASSO, por ende, se cambia la forma en que se penalizan los coeficientes.

En la Figura 5 se muestran los diferentes valores de los coeficientes para distintos valores de penalización. Podemos observar que los coeficientes tienden estabilizarse para pequeños valores de α. Utilizando nuevamente MSE como medida de ajuste, el valor de α tal que lo minimiza es aproximadamente 0.0027.



(a) Escala logarítmica

(b) Escala lineal

Figura 3: Coeficientes modelo Ridge en función del parámetro de penalización

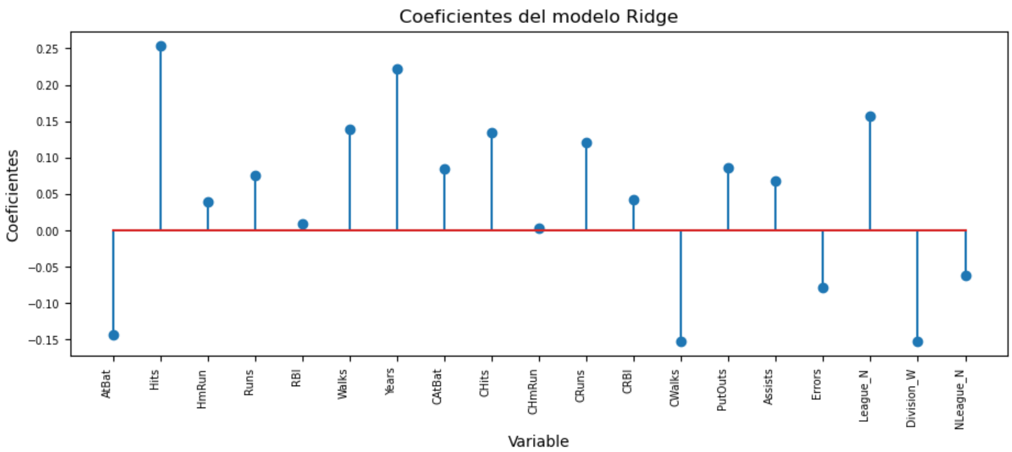
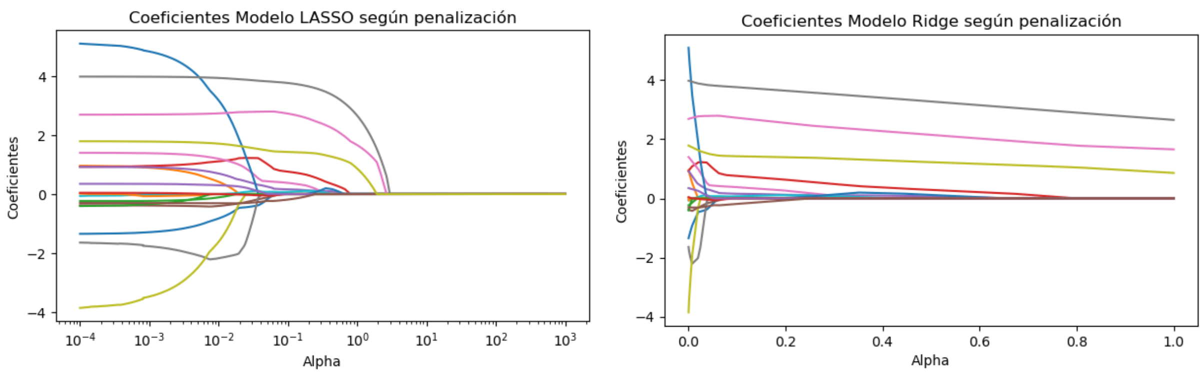


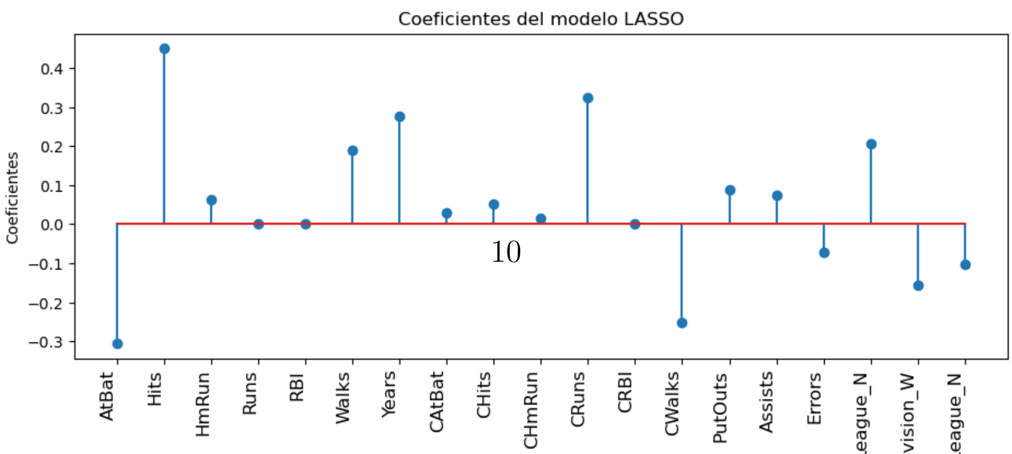
Figura 4: Correlaciones variables `Hitters`



(a) Escala logarítmica

(b) Escala lineal

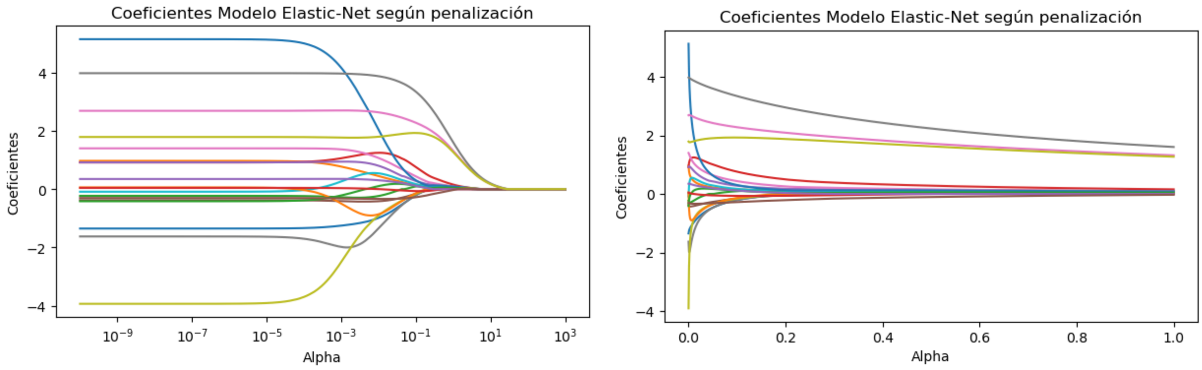
Figura 5: Coeficientes modelo LASSO en función del parámetro de penalización



En este caso, la Figura 6 nos dice que las variables que adquieren mayor importancia son `Hits`, `CRuns` y `Atbat` mientras que existe un mayor número de coeficientes poco relevantes, éstos son `Runs`, `RBI`, `ChumRun` y `CRBI`.

Regresión Elastic-Net

Al implementar el método Elastic-Net computacionalmente se obtuvo por validación cruzada, empleando 10 *k-folds*, que los parámetros que minimizan el error cuadrático medio es 0.1 para `l1_ratio` que corresponde al peso que se les otorga a la penalización LASSO, mientras que el 0.9 se considerará para el método Ridge. Por otro lado, el valor de α corresponde a 0.02 aproximadamente.



(a) Escala logarítmica (b) Escala lineal

Figura 7: Coeficientes modelo Elastic-Net en función del parámetro de penalización

Con el parámetro que entrega las ponderaciones para cada método se realizó la Figura 7 donde se aprecia que los errores, se observa que los valores se tienden a estabilizar en valores cercanos a 0.01. Utilizando validación cruzada notamos que el valor obtenido es 0.02. Posteriormente, se estudian los valores estimados de los parámetros.

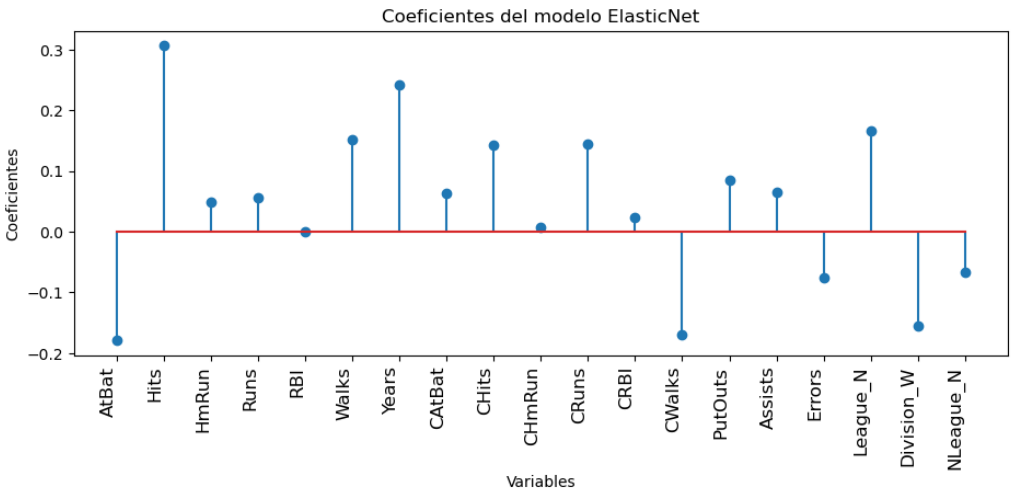


Figura 8: Estimaciones método Elastic-Net

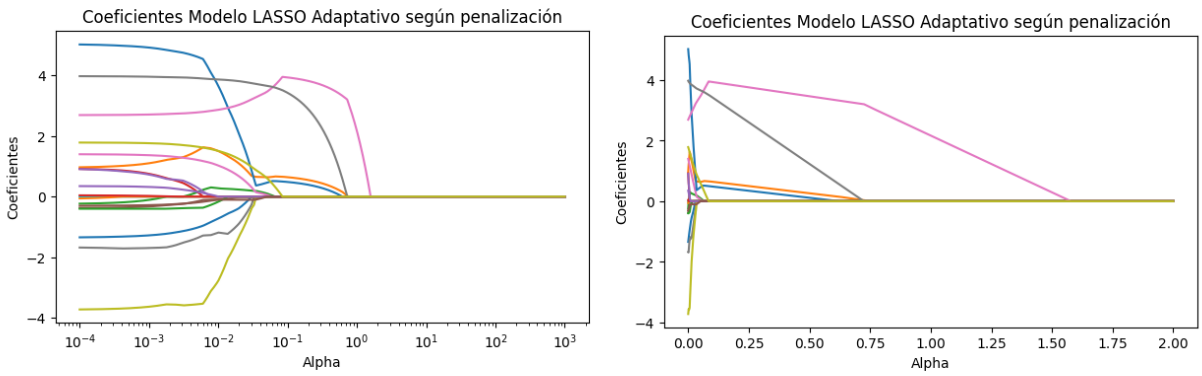
.

En la Figura 11 notmaos que los coeficientes más relevantes son `Hits` y `Years`, mietras que las menos importantes son `RBI`, `CHumRun` y `CRBI`.

Regresión LASSO adaptativo

Por último, se ajustó un modelo de regresión mediante penalización LASSO adaptativo, cuyos pesos corresponden a la función de los estimadores mínimos cuadrados $\hat{w} = 1/|\hat{\beta}_{EMCO}|^{\gamma}$, con $\gamma = 1$. Considerando una partición de 10 grupos (k-folds), el α que minimiza el error cuadrático medio es 0.0003.

En este caso, observando la figura 10 se concluye que las variables más importantes que ayudan a determinar el salario son `Hits`, `CRuns` y `AtBat`.



(a) Escala logarítmica

(b) Escala lineal

Figura 9: Coeficientes modelo LASSO Adaptativo en función del parámetro de penalización

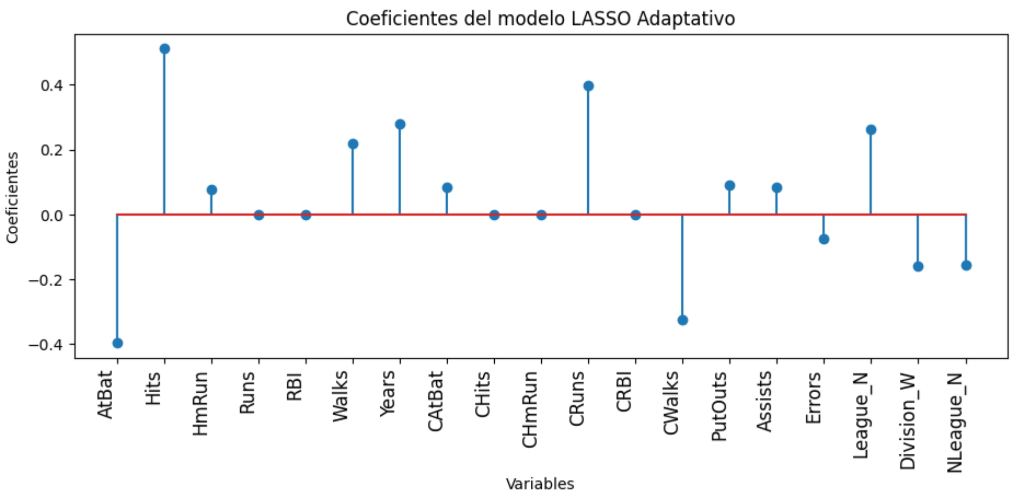


Figura 10: Estimaciones método LASSO Adaptativo

.

Variable	Ridge	Lasso	Lasso A.	Elastic Net
AtBat	-0.14	-0.31	-0.40	-0.18
Hits	0.25	0.45	0.51	0.31
HmRun	0.04	0.06	0.08	0.05
Runs	0.08	0.00	0.00	0.06
RBI	0.01	-0.00	0.00	0.00
Walks	0.14	0.19	0.22	0.15
Years	0.22	0.28	0.28	0.24
CAtBat	0.09	0.03	0.08	0.06
CHits	0.13	0.05	0.00	0.14
CHmRun	0.00	0.01	0.00	0.01
CRuns	0.12	0.32	0.40	0.14
CRBI	0.04	0.00	0.00	0.02
CWalks	-0.15	-0.25	-0.32	-0.17
PutOuts	0.09	0.09	0.09	0.08
Assists	0.07	0.07	0.08	0.07
Errors	-0.08	-0.07	-0.07	-0.08
League_N	0.16	0.21	0.26	0.17
Division_W	-0.15	-0.16	-0.16	-0.16
NLeague_N	-0.06	-0.10	-0.16	-0.07

Tabla 2: Resumen coeficientes según método

En resumen, en cada método se apreciaron valores estimados diferentes en los coeficientes. En ocasiones los coeficientes con mayor relevancia coincidían, mas no así su estimación. Por otro lado, también se observa algunas variables que tienden a tener una menor magnitud, siendo muy cercana o igual a cero. Lo anterior se debe a que cada método, si bien tienen planteamientos similares, es decir, agregar un término que penalice los coeficientes, éstos métodos difieren en la forma que se calculan. Así por ejemplo, por un lado en la regresión LASSO se utiliza la norma 1 que permite que los coeficientes sean cero, por el otro, la regresión Ridge, por su norma 2, usualmente no permite que los coeficientes sean iguales a cero. Del mismo modo, Elastic-Net corresponde a una ponderación de ambos métodos.

Así, las variables con mayor grado de importancia en cada método para predecir el salario del jugador son el número de hits (`Hits`), el número de turnos que tomó para batear (`AtBat`) y los años (`Years`) que lleva jugando también permanece dentro de los factores más influyentes.

Solución (b)

Para ajustar cada uno de los modelos se considero un set de entrenamiento, validación y testeo, utilizando el 60 %, 20 % y 20 % del total de datos respectivamente. Donde se ajustan los modelo utilizando el test de entrenamiento, se calibran los parámetros con el set de validación y se realizan las predicciones con el set de testeo, y con éstos últimos se calcula el error en la predicción.

Posterior a la implementación de cada uno de los modelos según lo solicitado, se determina el promedio del error cuadrático medio (MSE) de cada método.

Método	Promedio MSE
Mínimos Cuadrado	0.452
Regresión Ridge	0.398
LASSO	0.398
Elastic-Net	0.381
LASSO Adaptativo	0.45

Tabla 3: Error cuadrático medio para cada método

La información disponible en la Tabla 3 muestra que los métodos poseen una mejor capacidad predictiva en comparación con el método de mínimos cuadrados, además se aprecia que la diferencia no son comparativamente sustanciales. También, cabe señalar que el error de LASSO Adaptative es bastante similar al de mínimos cuadrados, esto de puede deber a la aleatoriedad que se dividió el conjunto de datos.

Preguntas (iv-v)

Se puede apreciar que los métodos implementados presentan un menor error en la predicción de los datos, principalmente el modelo Elastic-Net presenta un menor MSE promedio, sin embargo, esta cifra no es sustancialmente mayor, en comparación con los métodos Ridge y LASSO.

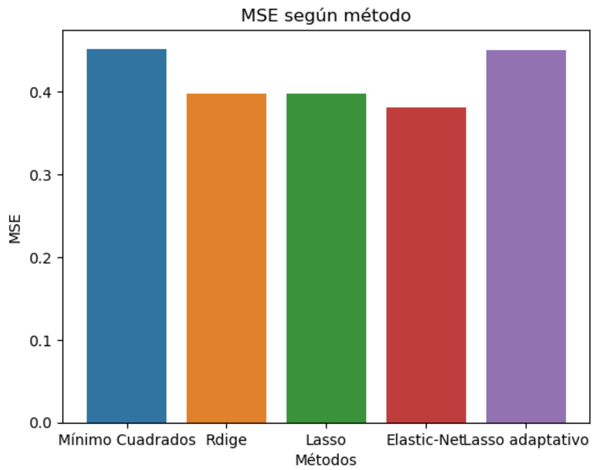


Figura 11: Estimaciones método Elastic-Net

.

Estos metodologías presentan ventajas al método de mínimos cuadrados, ya que como se observa en la Figura 2 existe variables altamente correlacionadas, lo que haría que las columnas de la matriz de diseño sean similares a columnas linealmente dependientes y, por consiguiente, produzca problemas para invertir la matriz $X^T X$, lo que a su vez podría afectar negativamente en las estimaciones y en la varianza de estas mismas. Por el contrario, los métodos implementados, intentar remediar esta situación considerando una penalización. Lo anterior, por lo general, muestra menor variabilidad en las estimaciones, aunque éstas sean sesgadas. Asimismo, estos métodos fueron ideados para lidiar con conjuntos de datos que posean una gran cantidad de predictores, lo que por el contrario, perjudica el método de mínimos cuadrados, ya que es más probable que más de algunos predictores estén correlacionados. Posteriormente, se muestran los resultados de las estimaciones mediante gráficos.

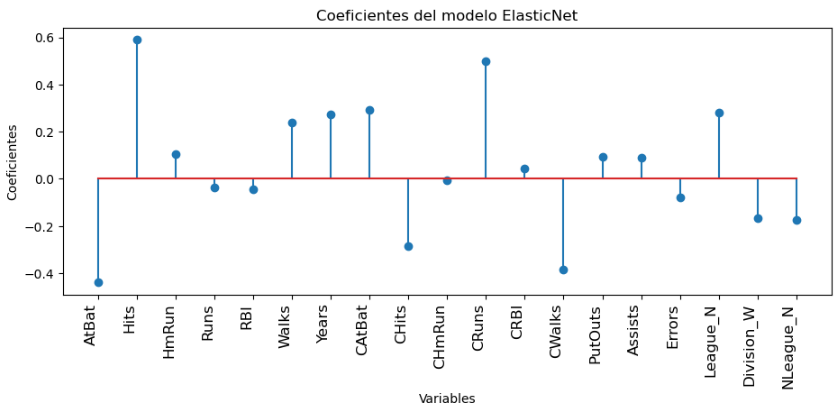


Figura 12: Estimaciones método Mínimo Cuadrados utilizando parámetro mejor calibrado

.

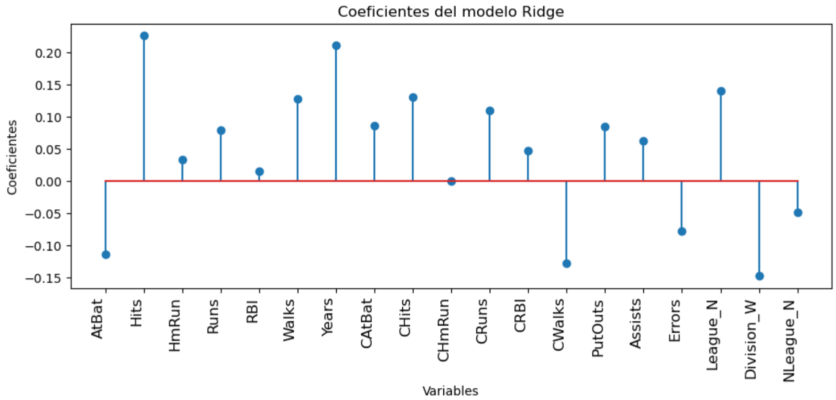


Figura 13: Estimaciones método RidgeElastic-Net utilizando parámetro mejor calibrado

.

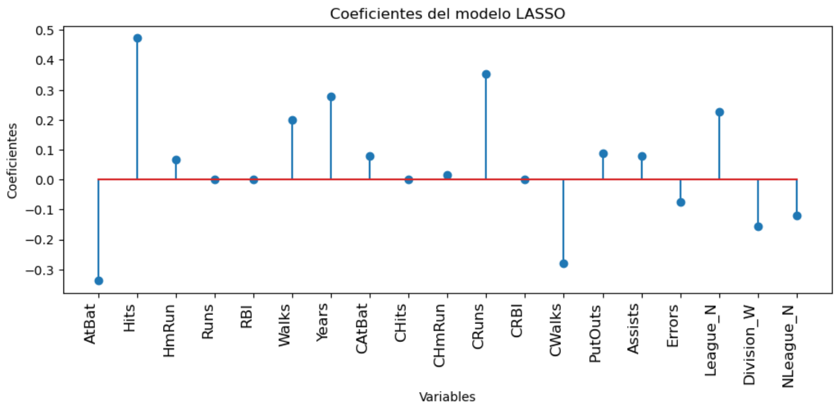


Figura 14: Estimaciones método LASSO utilizando parámetro mejor calibrado

.

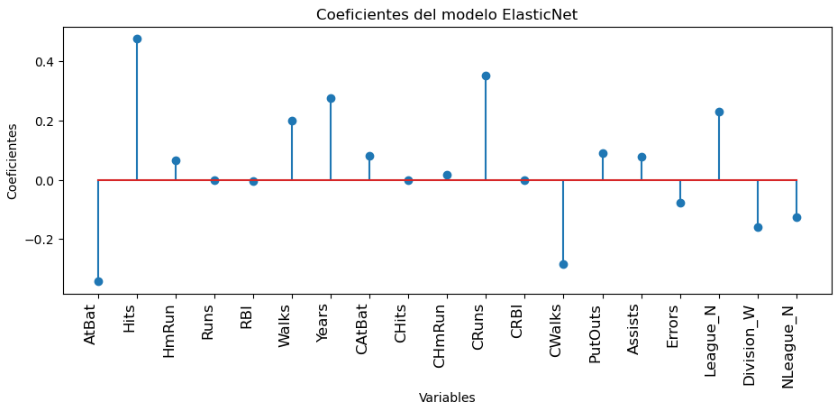


Figura 15: Estimaciones método Elastic-Net utilizando parámetro mejor calibrado

.

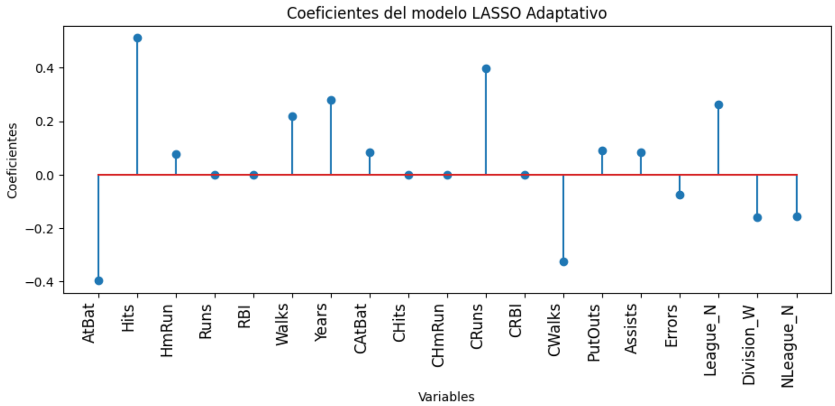


Figura 16: Estimaciones método Adaptative Lasso utilizando parámetro mejor calibrado

.

Se puede observar en las Figuras 12 - 16 que los métodos tienden a penalizar los mismos parámetros, salvo la regresión Ridge. Éstos coinciden con valores de estimaciones pequeños en el método de mínimos cuadrados, es decir, que posterior a esos métodos las estimaciones son aún más cercanas a cero, éstas variables corresponden a `Runs`, `RBI`, `CHits`, `CHumRun` y `CRBI`. Por el contrario, `Hits` y `CRuns` son variables que en la mayoría de los métodos, el valor de sus estimaciones es mayor que el del resto.

Finalmente, podemos notar que en comparación a utilizar todos los datos (y solo validación cruzada mediante las funciones `metodoCV`) notamos que al utilizar validación cruzada, al seleccionar los "mejores predictores"los resultados fueron similares entre los métodos.