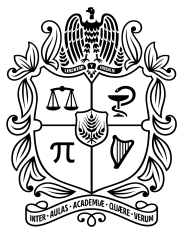
Clase 5 - Módulo 2: Introducción a la analítica

Mauricio Alejandro Mazo Lopera

Universidad Nacional de Colombia
Facultad de Ciencias
Escuela de Estadística
Medellín



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Métodos de contracción (Shrinkage)

Se busca principalmente **reducir la varianza de las estimadores de los parámetros**.

Métodos de contracción (Shrinkage)

Se busca principalmente **reducir la varianza de las estimadores de los parámetros**.

Recuerde que:

$$RSS = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$$

Métodos de contracción (Shrinkage)

Se busca principalmente **reducir la varianza de las estimadores de los parámetros**.

Recuerde que:

$$RSS = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$$

Los dos métodos de **contracción** o **regularización** son:

Ridge **Lasso**

Ridge

Se busca minimizar

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

o lo que es equivalente:

$$\min_{\beta} \{RSS\} \text{ sujeto a}$$

$$\sum_{j=1}^{p} \beta_j^2 \leq s$$

Lasso

Se busca minimizar

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

o lo que es equivalente:

$$\min_{\beta} \{RSS\} \text{ sujeto a}$$

$$\sum_{j=1}^{p} |\beta_j| \leq s$$

Observaciones sobre Ridge y Lasso

- λ es conocido como parámetro de calibración y se puede obtener con validación cruzada, por ejemplo.

Observaciones sobre Ridge y Lasso

- λ es conocido como parámetro de calibración y se puede obtener con validación cruzada, por ejemplo.
- Denotemos por $\widehat{\beta}$, $\widehat{\beta}_R$ y $\widehat{\beta}_L$ los vectores de parámetros estimados por mínimos cuadrados ordinarios, por Ridge y por Lasso, respectivamente.

Observaciones sobre Ridge y Lasso

MLS **Ridge**

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \qquad \hat{\beta}_R = (\boldsymbol{X}'\boldsymbol{X} + \lambda \boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

- Cuando $\lambda \longrightarrow 0$ se tiene que

$$\hat{\beta}_R \longrightarrow \hat{\beta} \quad \text{y} \quad \hat{\beta}_L \longrightarrow \hat{\beta}$$

Observaciones sobre Ridge y Lasso

MLS **Ridge**

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \qquad \hat{\beta}_R = (\boldsymbol{X}'\boldsymbol{X} + \lambda \boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

- Cuando $\lambda \longrightarrow 0$ se tiene que

$$\hat{\beta}_R \longrightarrow \hat{\beta} \quad \text{y} \quad \hat{\beta}_L \longrightarrow \hat{\beta}$$

- Cuando $\lambda \longrightarrow \infty$ se tiene que

$$\hat{\beta}_R \longrightarrow \mathbf{0} \quad \text{y} \quad \hat{\beta}_L \longrightarrow \mathbf{0}$$

Observaciones sobre Ridge y Lasso

- Dada la sensibilidad que tienen estos métodos con respecto a la escala de las covariables, se recomienda estandarizar antes de aplicarlos, es decir, definir

$$\tilde{x}_{ij} = \frac{x_{ij}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2}}$$

Observaciones sobre Ridge y Lasso

- Dada la sensibilidad que tienen estos métodos con respecto a la escala de las covariables, se recomienda estandarizar antes de aplicarlos, es decir, definir

$$\tilde{x}_{ij} = \frac{x_{ij}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2}}$$

- Cuando λ aumenta, la flexibilidad del modelo disminuye, lo cual lleva a que la varianza de los betas estimados disminuya, pero aumente el sesgo.

Observaciones sobre Ridge y Lasso

- Cuando $p \approx n$ o $p > n$, la varianza de los betas estimados aumenta considerablemente cuando se usa mínimos cuadrados. Por tanto, se recomendaría el uso del proceso de regularización.

Observaciones sobre Ridge y Lasso

- Cuando $p \approx n$ o $p > n$, la varianza de los betas estimados aumenta considerablemente cuando se usa mínimos cuadrados. Por tanto, se recomendaría el uso del proceso de regularización.
- A diferencia de Ridge, Lasso sí lleva a estimaciones de los parámetros exactamente iguales a cero, lo cual mejora las interpretaciones, ya que permite excluir variables no significativas.

Observaciones sobre Ridge y Lasso

- Cuando $p \approx n$ o $p > n$, la varianza de los betas estimados aumenta considerablemente cuando se usa mínimos cuadrados. Por tanto, se recomendaría el uso del proceso de regularización.
- A diferencia de Ridge, Lasso sí lleva a estimaciones de los parámetros exactamente iguales a cero, lo cual mejora las interpretaciones, ya que permite excluir variables no significativas.
- No hay una ventaja clara de un método sobre otro, esto depende de los datos y por tanto, es posible aplicar validación cruzada para seleccionar entre ambos.

Selección del parámetro de calibración λ

Para poder aplicar Ridge o Lasso es necesario encontrar el valor de penalización λ (también conocido como parámetro de calibración). El método más utilizado consiste en:

1. Seleccionar un conjunto de valores de λ en un intervalo abierto (generalmente cero es el límite inferior).
2. Usar validación cruzada para cada valor de λ y estimar el error de validación cruzada.
3. Escoger el λ que produzca el menor error.

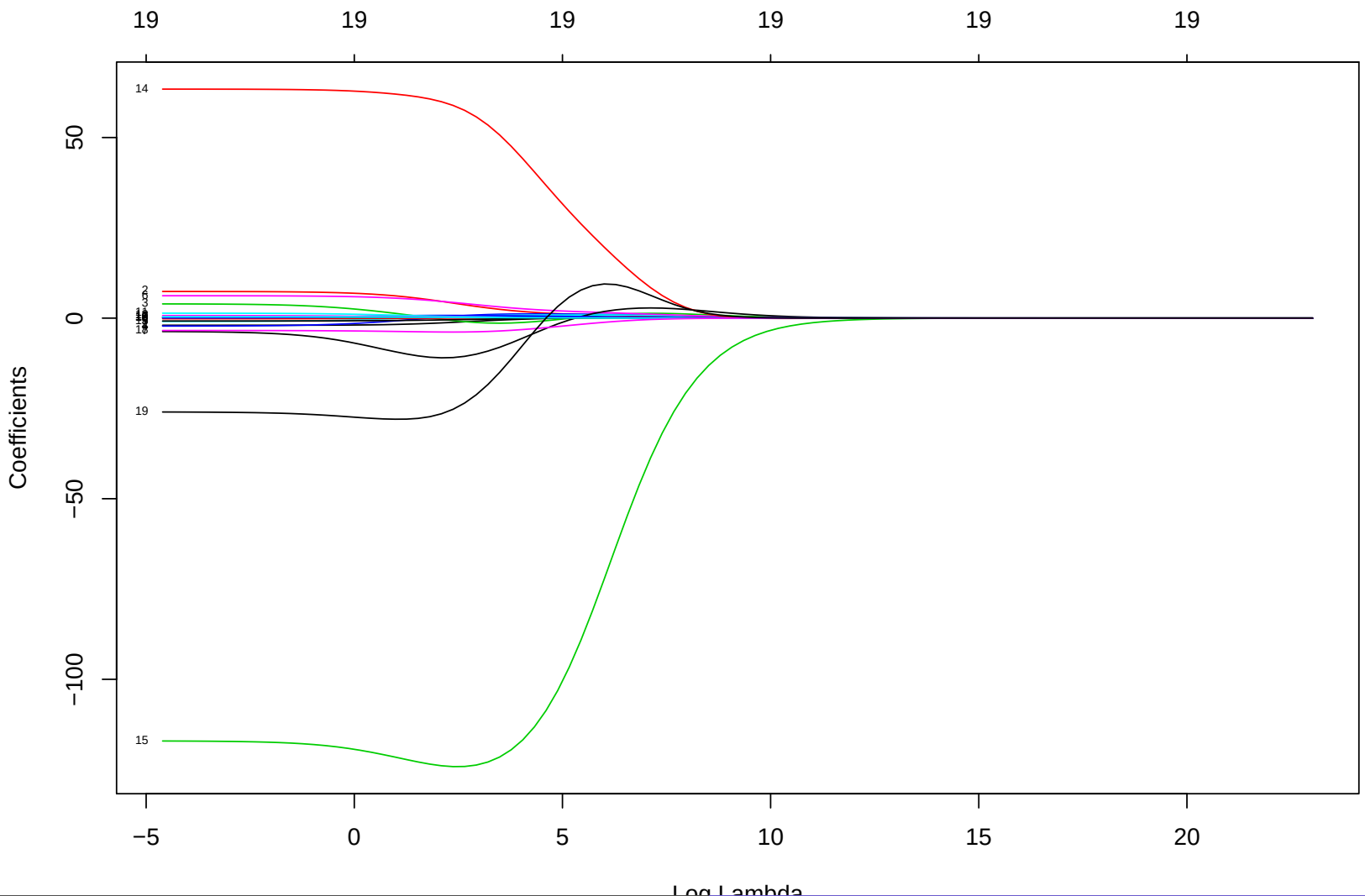
Trabajando con R: Ridge regression

```
require(ISLR)
require(glmnet)
Hitters<-na.omit(Hitters)
x<-model.matrix(Salary~.,Hitters)[,-1]
y<-Hitters$Salary

gridz<-10^seq(-2,10, length=100)
ridge.mod<-glmnet(x,y,alpha=0, lambda=gridz)

dim(coef(ridge.mod))
```

```
## [1]  20 100
```

Trabajando con R: Ridge regression

```
plot(ridge.mod, xvar="lambda", label=TRUE)
```

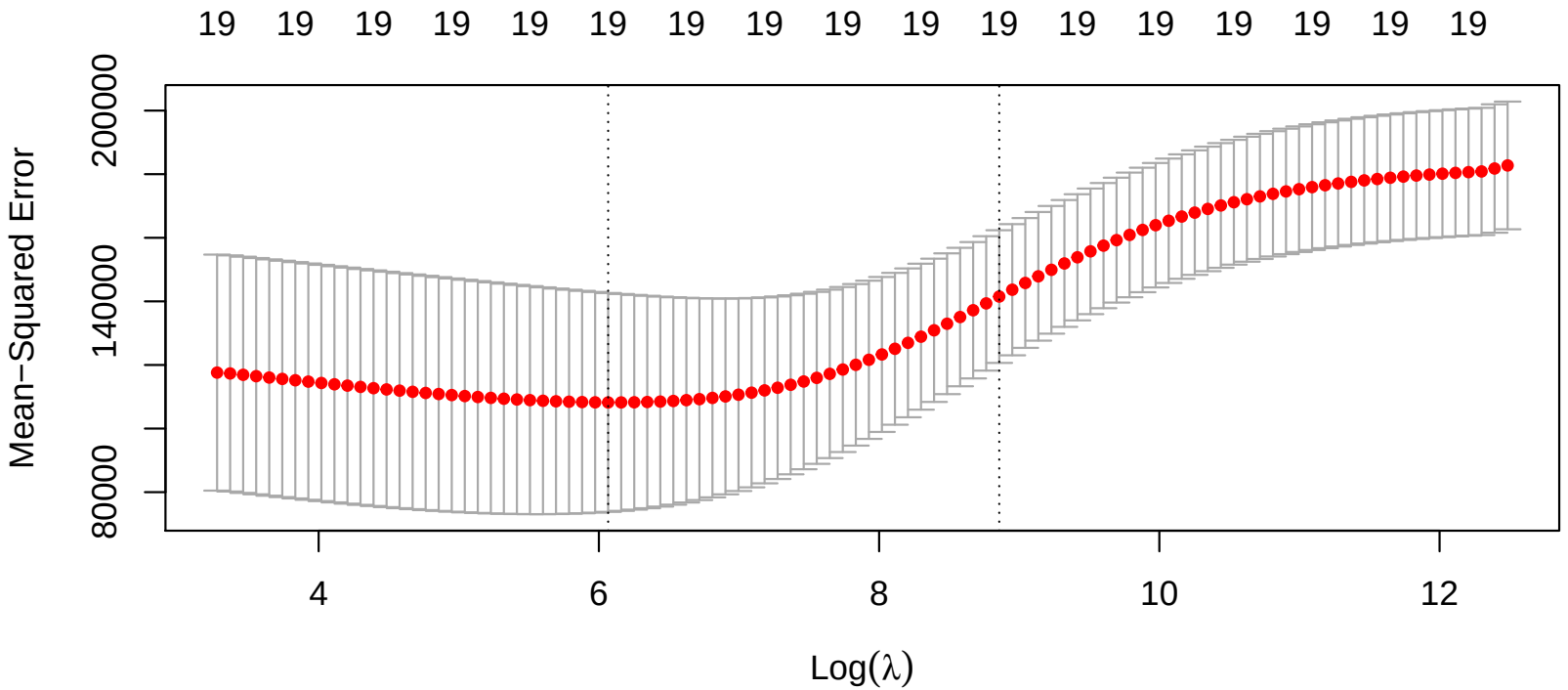


Trabajando con R: Ridge regression

```
cedula<-1
set.seed(cedula)
train<-sample(1: nrow(x), nrow(x)/2)
test<- -train
y.test<-y[test]
cv.out<-cv.glmnet(x[train,],y[train],alpha=0)
```

Trabajando con R: Ridge regression

```
plot(cv.out)
```




```
bestlam<-cv.out$lambda.min
bestlam
```

```
## [1] 431.0623
```

Trabajando con R: Ridge regression

```
ridge.pred<-predict(ridge.mod, s=bestlam,newx=x[test,])
mean((ridge.pred-y.test)^2)
```

```
## [1] 120779.4
```

```
out<-glmnet (x,y,alpha=0)
predict(out,type="coefficients",s=bestlam)[1:20,]
```

```
##  (Intercept)        AtBat         Hits        HmRun         Runs          RB
## 24.88203059   0.09571573   0.77893014   0.85488975   1.02412478   0.8757581
##        Walks        Years       CAtBat        CHits       CHmRun         CRun
##  1.51400163   1.93680145   0.01128321   0.05331300   0.38052103   0.1065181
##         CRBI       CWalks      LeagueN    DivisionW      PutOuts      Assist
##  0.11215778   0.06298811 18.94995828 -70.27086691   0.14893846   0.0232910
##       Errors   NewLeagueN
## -1.09701572   9.45875165
```

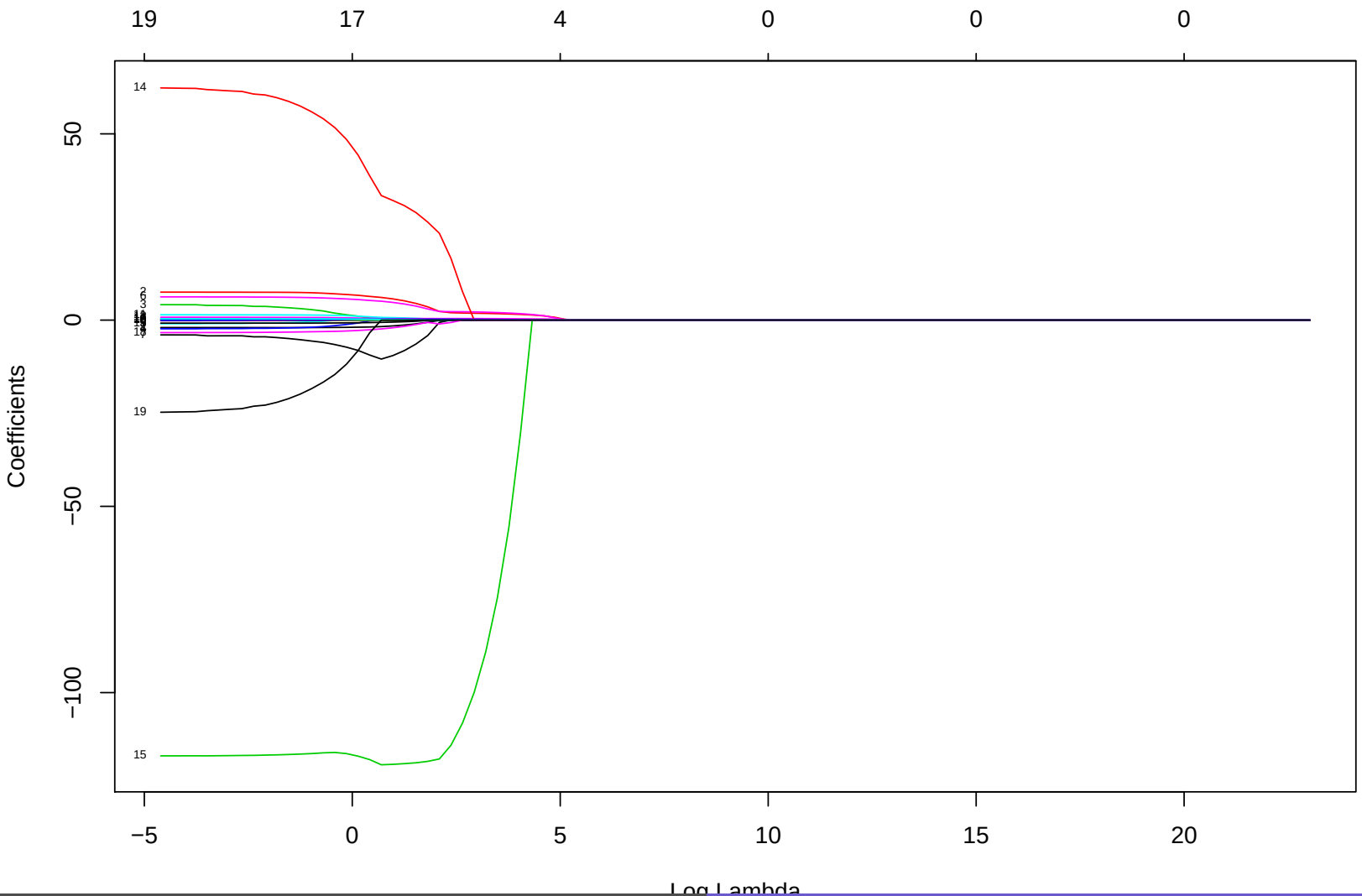
Trabajando con R: Lasso regression

```
require(ISLR)
require(glmnet)
Hitters<-na.omit(Hitters)
x<-model.matrix(Salary~.,Hitters)[,-1]
y<-Hitters$Salary

gridz<-10^seq(-2,10, length=100)
lasso.mod<-glmnet(x,y,alpha=1, lambda=gridz)

dim(coef(lasso.mod))
```

```
## [1]  20 100
```

Trabajando con R: Ridge regression

```
plot(lasso.mod, xvar="lambda", label=TRUE)
```

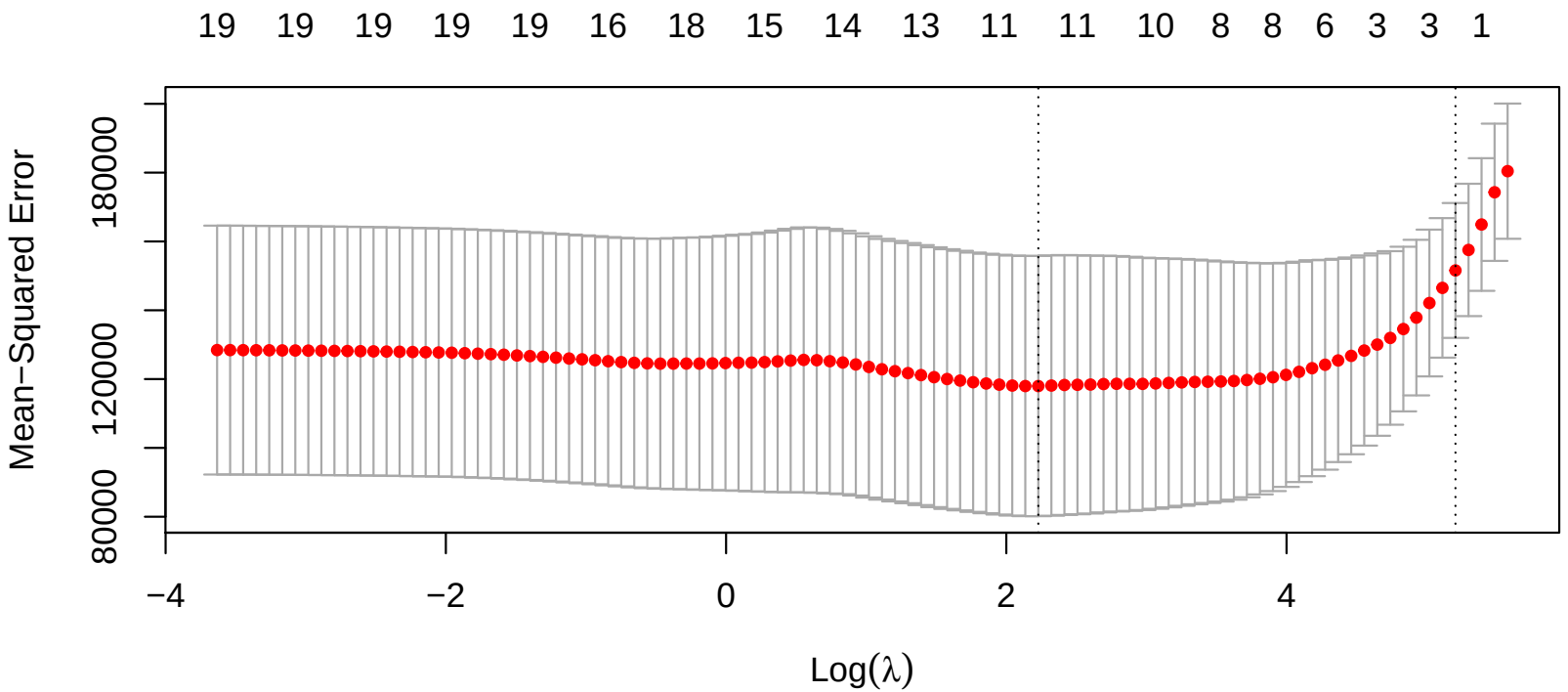


Trabajando con R: Lasso regression

```
cedula<-1
set.seed(cedula)
train<-sample(1: nrow(x), nrow(x)/2)
test<- -train
y.test<-y[test]
cv.out<-cv.glmnet(x[train,],y[train],alpha=1)
```

Trabajando con R: Lasso regression

```
plot(cv.out)
```




```
bestlam<-cv.out$lambda.min
bestlam
```

```
## [1] 9.286955
```

Trabajando con R: Lasso regression

```
lasso.pred<-predict(lasso.mod, s=bestlam,newx=x[test,])
mean((lasso.pred-y.test)^2)
```

```
## [1] 111754.2
```

```
out<-glmnet (x,y,alpha=1)
lasso.coef<-predict(out,type="coefficients",s=bestlam)[1:20,]
lasso.coef
```

```
##   (Intercept)         AtBat          Hits         HmRun          Runs
##   -3.04787648    0.00000000    2.02551572    0.00000000    0.00000000
##           RBI         Walks         Years        CAtBat         CHits
##    0.00000000    2.26853781    0.00000000    0.00000000    0.00000000
##        CHmRun         CRuns          CRBI        CWalks       LeagueN
##    0.01647106    0.21177390    0.41944632    0.00000000   20.48456543
##     DivisionW       PutOuts       Assists        Errors    NewLeagueN
## -116.59062078    0.23718459    0.00000000   -0.94739923    0.00000000
```

Trabajando con R: Lasso regression

```
lasso.coef[lasso.coef!=0]
```

```
##   (Intercept)          Hits         Walks      CHmRun         CRuns
##  -3.04787648    2.02551572    2.26853781   0.01647106   0.21177390
##          CRBI       LeagueN     DivisionW     PutOuts       Errors
##   0.41944632   20.48456543 -116.59062078   0.23718459  -0.94739923
```

Actividad:

Utilice las técnicas ridge y lasso para regularizar las bases de datos **BASE_DATOS_1** y **BASE_DATOS_2**. Según estas técnicas, ¿cuáles variables aparentemente muestran no ser relevantes para explicar la variable aleatoria Y?