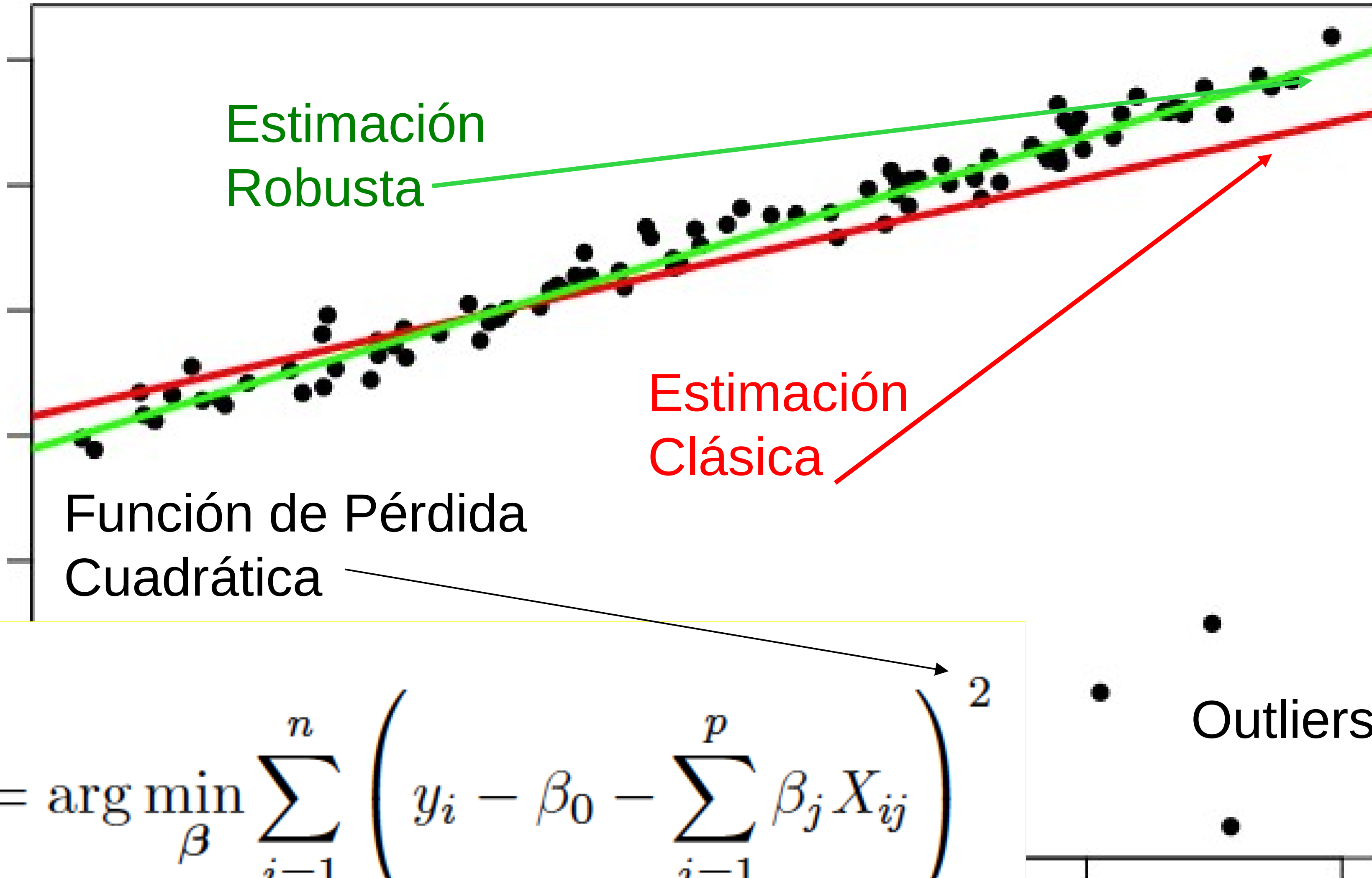
Robustez

Protección ante datos atípicos (outliers)

La Importancia de la Robustez

- La información posee frecuentemente datos atípicos (outliers):
  - Mediciones erroneas
  - Mediciones extremas
- Los datos atípicos pueden afectar fuertemente el resultado de la aplicaciión de los métodos clásicos.
- Puede ser de interes la detección de los datos atípicos (Novelty detection).

El problema



$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij} \right)^2$$

Soluciones en Regresión

Estimador L1

Valor absoluto

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left| y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij} \right|$$

Estimador LMS

Mediana

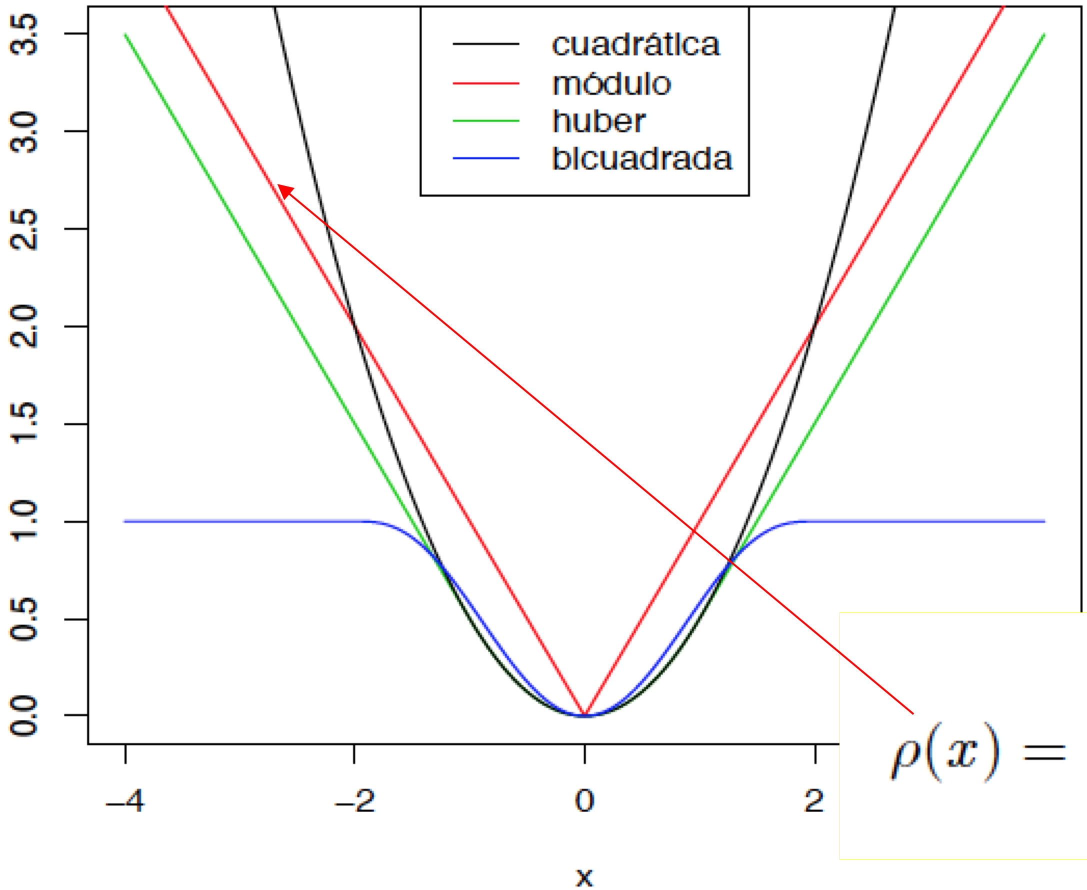
$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \text{med} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij} \right)^2 \right\}$$

Técnicas más Eficientes

M-Estimadores de Regresión

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij}}{\hat{\sigma}}\right)$$

Estimador robusto de escala de los errores



Función de Huber

$$\rho(x) = \begin{cases} x^2 & \text{si} \quad x \leq k \\ 2k\,|x| - k^2 & \text{si} \quad x > k \end{cases}$$

Ecuaciones de estimación

Derivando se obtienen las ecuaciones des estimación para el estimador de mínimos cuadrados (ecuaciones normales):

$$\boxed{\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j X_{ij}\right)\mathbf{X}_i = 0}$$

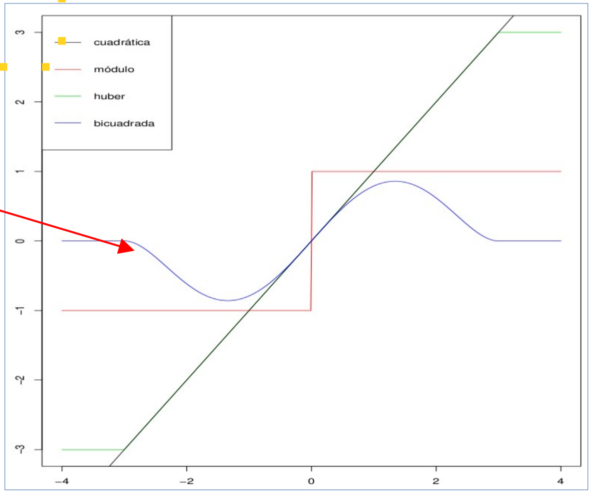
Los parámetros estimados satisfacen esta ecuación

para M-estimadores :

$$\boxed{\sum_{i=1}^{n}\psi\left(\frac{y_i - \beta_0 - \sum_{j=1}^{p}\beta_j X_{ij}}{\hat{\sigma}}\right)\mathbf{X}_i = 0}$$

donde $\psi = \rho'$

Derivada de la Función de Pérdida



La ecuación

$$\sum_{i=1}^{n} \psi\left(\frac{y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij}}{\hat{\sigma}}\right) \mathbf{X}_i = 0$$

Nueva ecaución de estimación con PESOS

es equivalente a

$$\sum_{i=1}^{n} \frac{\psi\left(\frac{y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij}}{\hat{\sigma}}\right)}{y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij}} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij}\right) \mathbf{X}_i = 0,$$

o sea

$$\boxed{\sum_{i=1}^{n} W_i \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij}\right) \mathbf{X}_i = 0,}$$

Pesos basados en la ATIPICIDAD de las obs.

con

$$W_i = \frac{\psi\left(\frac{y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij}}{\hat{\sigma}}\right)}{y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij}}$$

Es la ecuación de un estimador de mínimos cuadrados pesados con pesos desconocidos. Se resuelve por un métodos iterativos: IRWLS.

Implementación en R

- Rlm (MASS)
- rq (quantreg)
- lmRob (robust)
- lmrob (robustbase)

https://cran.r-project.org/web/views/Robust.html

CRAN Task View: Robust Statistical Methods

Regularización (Shrinkage)

Regresión con Penalización
Ridge y Lasso

Regularización, para que se usa ?

- En general, es una técnica para prevenir el Overfitting, mediante la penalización de la complejidad del modelo.
- Controla los desbalances entre el número de observaciones y el número de variables.
- Como método automático de selección de variables (Lasso).
- Reduce la alta variabilidad observada en situaciones de multicolinealidad (Ridge).
- Cuidado ! Reduce la Varianza pero Aumenta el Sesgo de los estimadores !
- Cuidado ! No son equivariantes por cambios de escala ! Hay que estandarizar las variables !

Regresión Ridge

Parámetro de Penalización

Penalización cuadrática (L2)

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

Falta de Ajuste

Penalización, encoge los coeficientes hacia el CERO (Shrinkage)

Propiedades del Estimador Ridge

$$\boxed{\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}}$$

Distinto de CERO y creciente (en modulo) con Lambda

Decreciente con Lambda

$$\mathrm{Bias}(\widehat{\boldsymbol{\beta}}) = -\lambda\mathbf{W}\boldsymbol{\beta}$$

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \sigma^2\mathbf{W}\mathbf{X}^T\mathbf{X}\mathbf{W}$$

Matriz Inversible (SIEMPRE)

$$\boxed{\mathbf{W} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}}$$

Resultado Sorprendente !!!

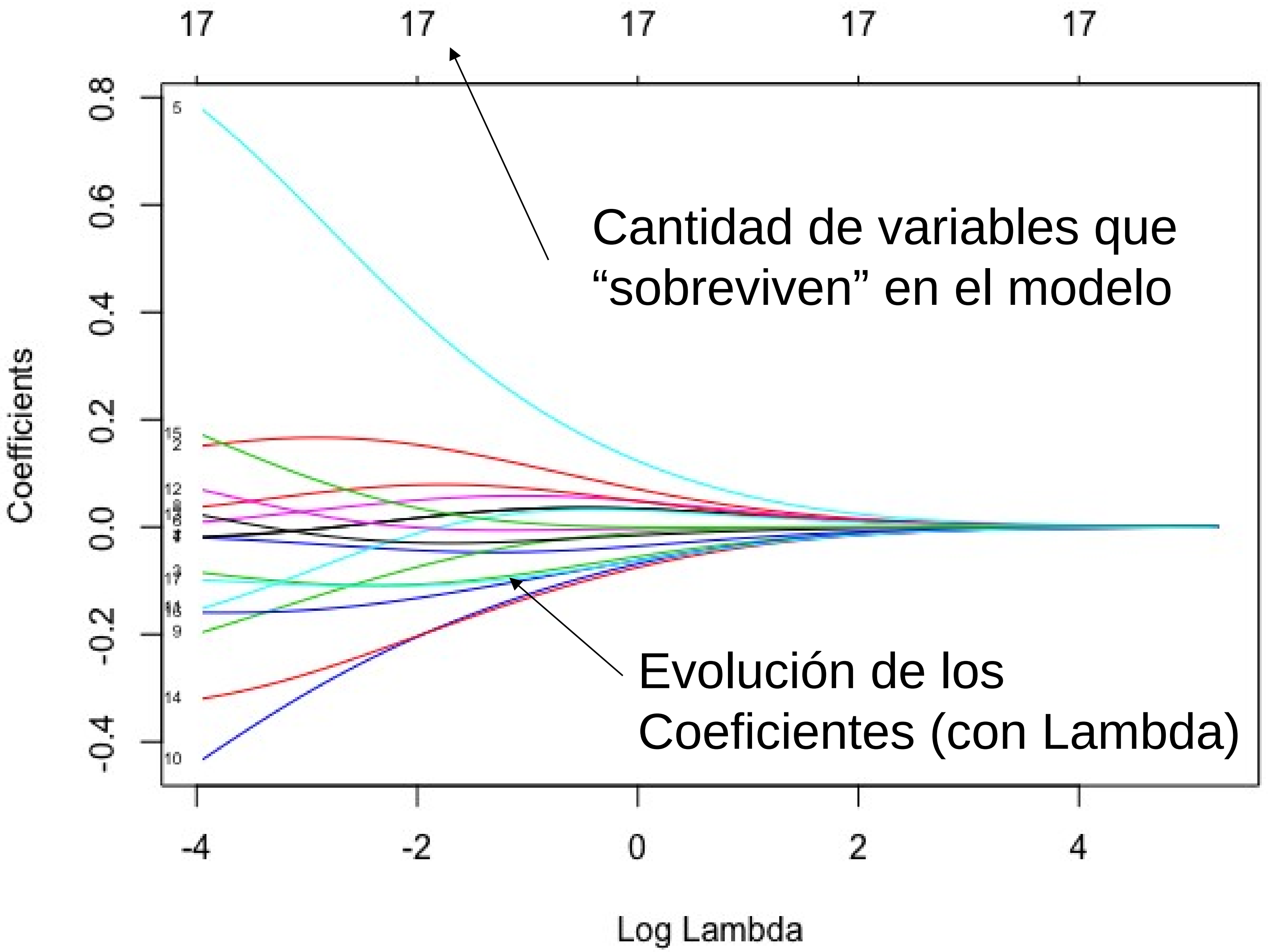
- Siempre existe un valor del parámetro de peanalización tal que el MSE de Ridge es estrictamente menor que el MSE de OLS.

$$\exists \lambda > 0 \qquad MSE_{\hat{\beta}_\lambda}^{RIDGE} < MSE_{\beta}^{OLS}$$

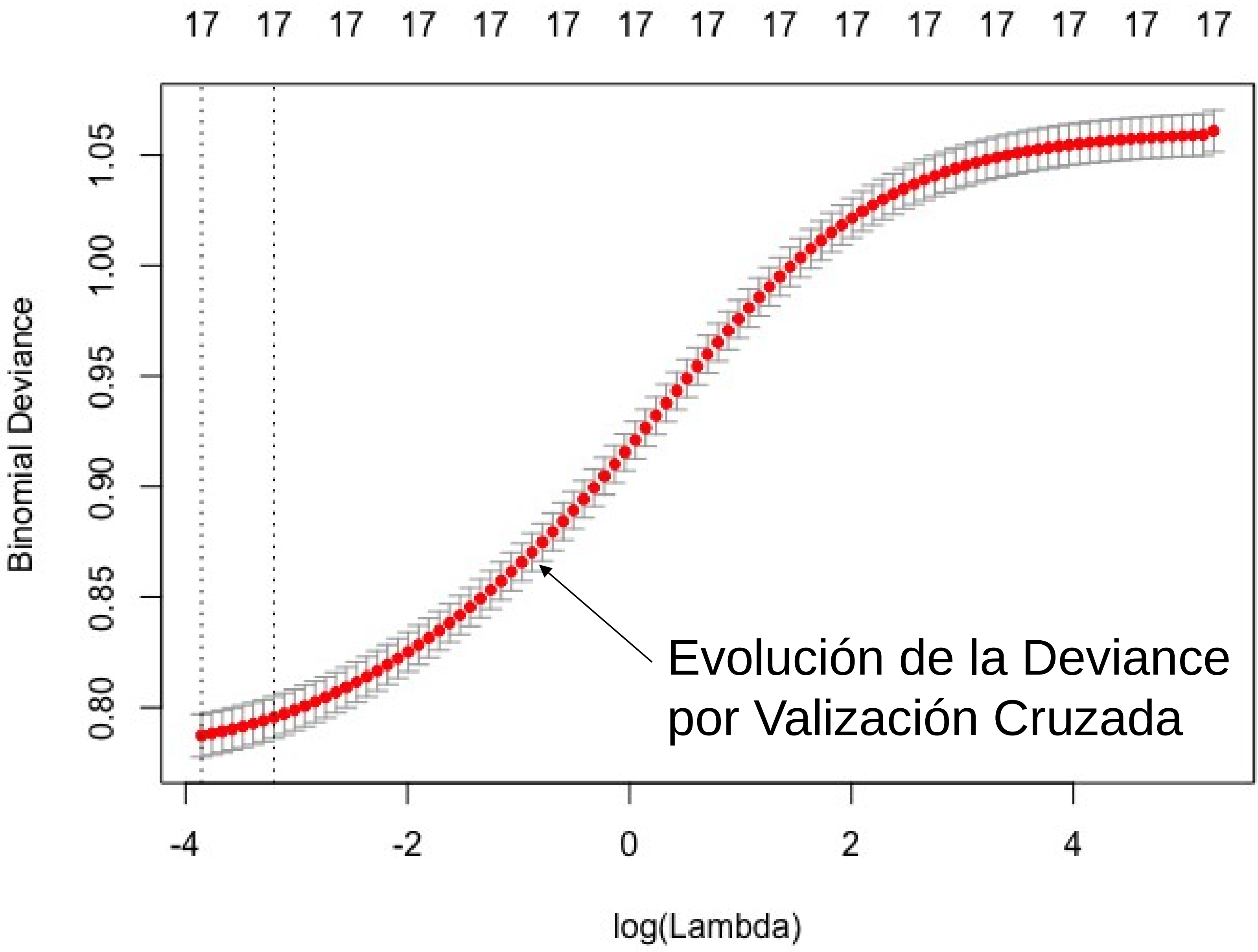
- Como se elige el Lambda óptimo ??????Validación Cruzada

Ejemplo de Ridge Regression

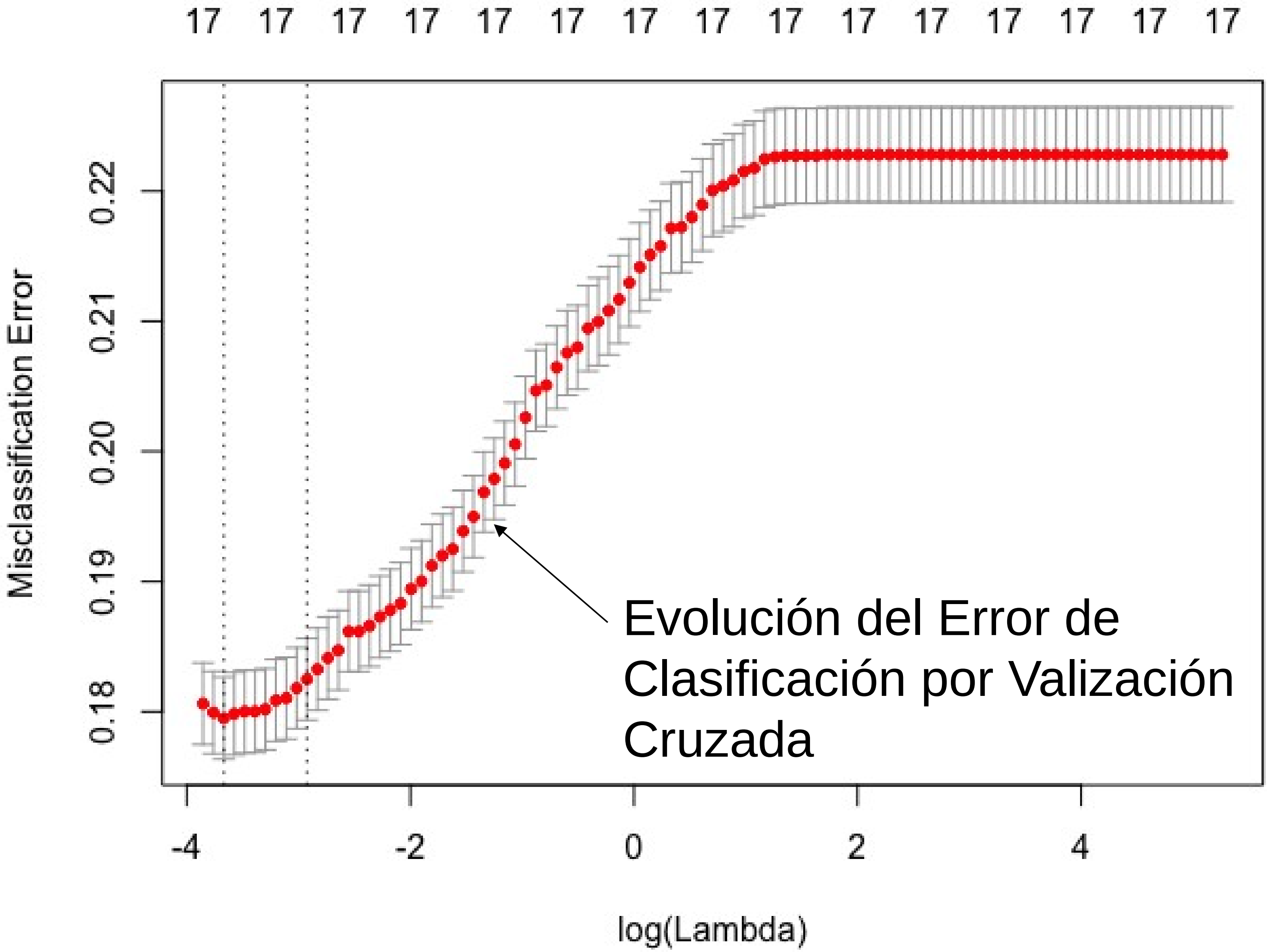
Predicción de Lluvia



Elección de Lambda (Deviance)



Elección del Lambda (Error Clsificación)



Lasso Regression

- Puede usarse como método de selección de variables

Parámetro de Penalización

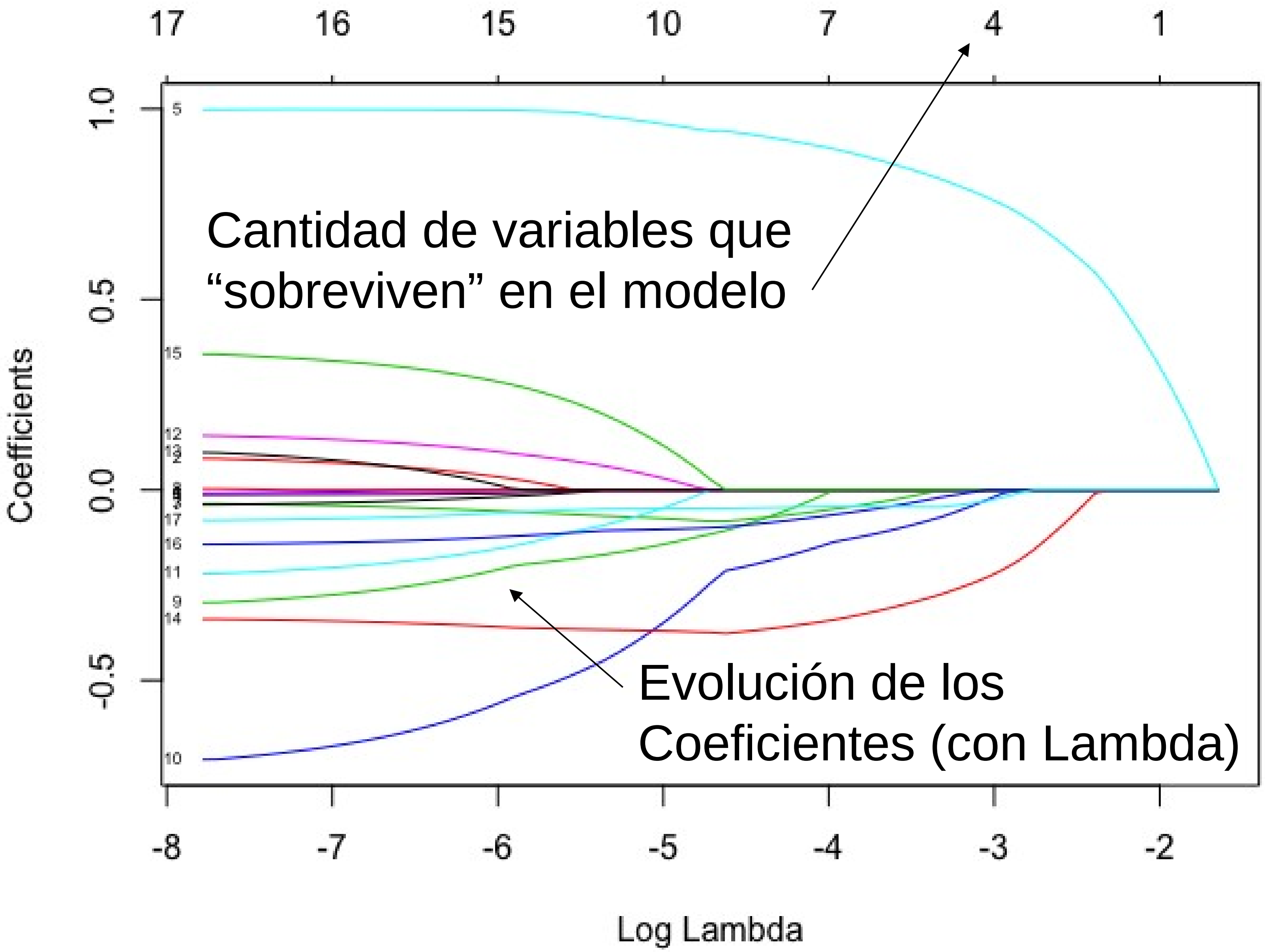
Penalización lineal (L1)

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

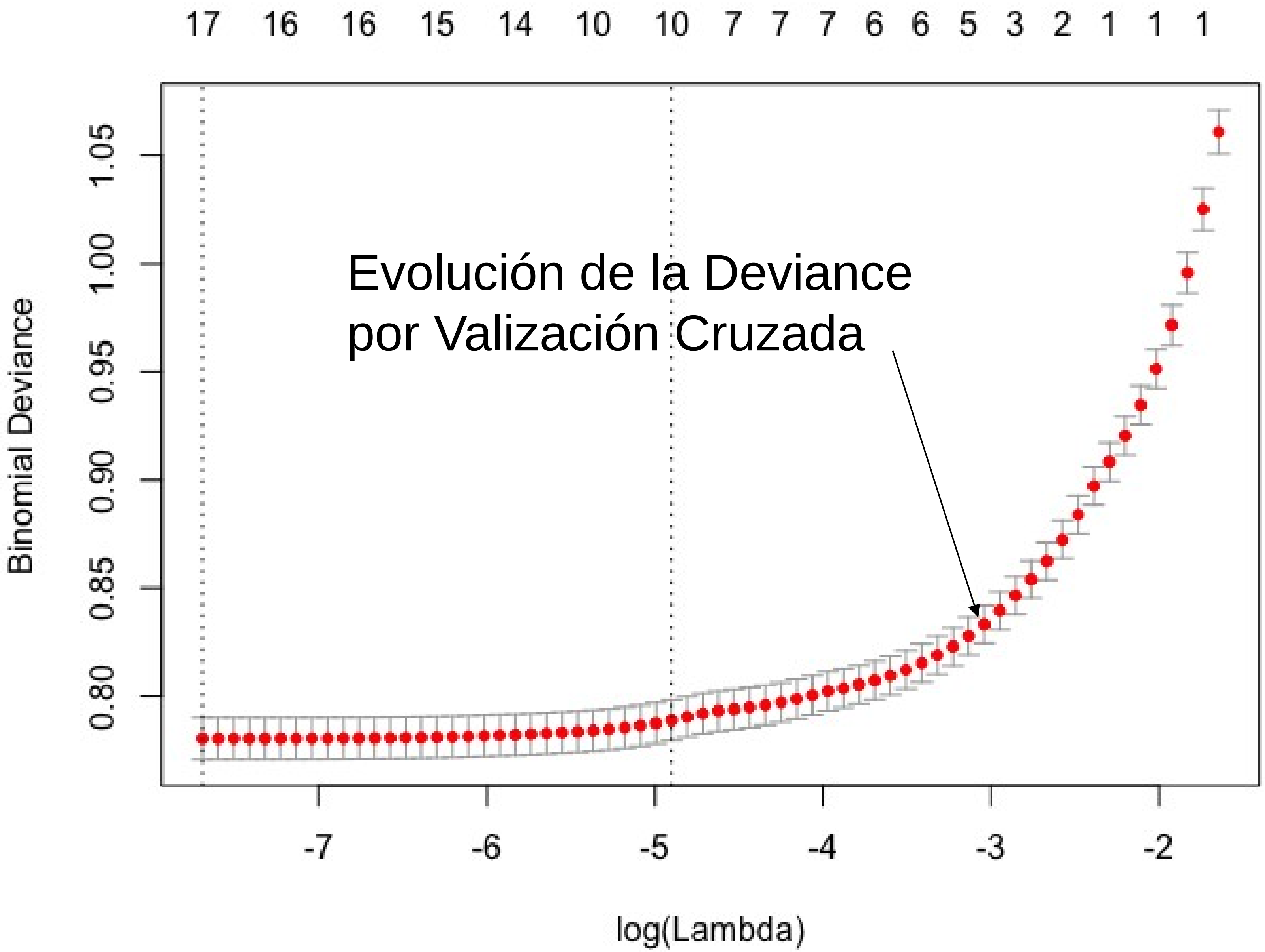
Falta de Ajuste

Penalización, encoge los coeficientes hacia el CERO. Convierta a algunos coeficientes en CERO

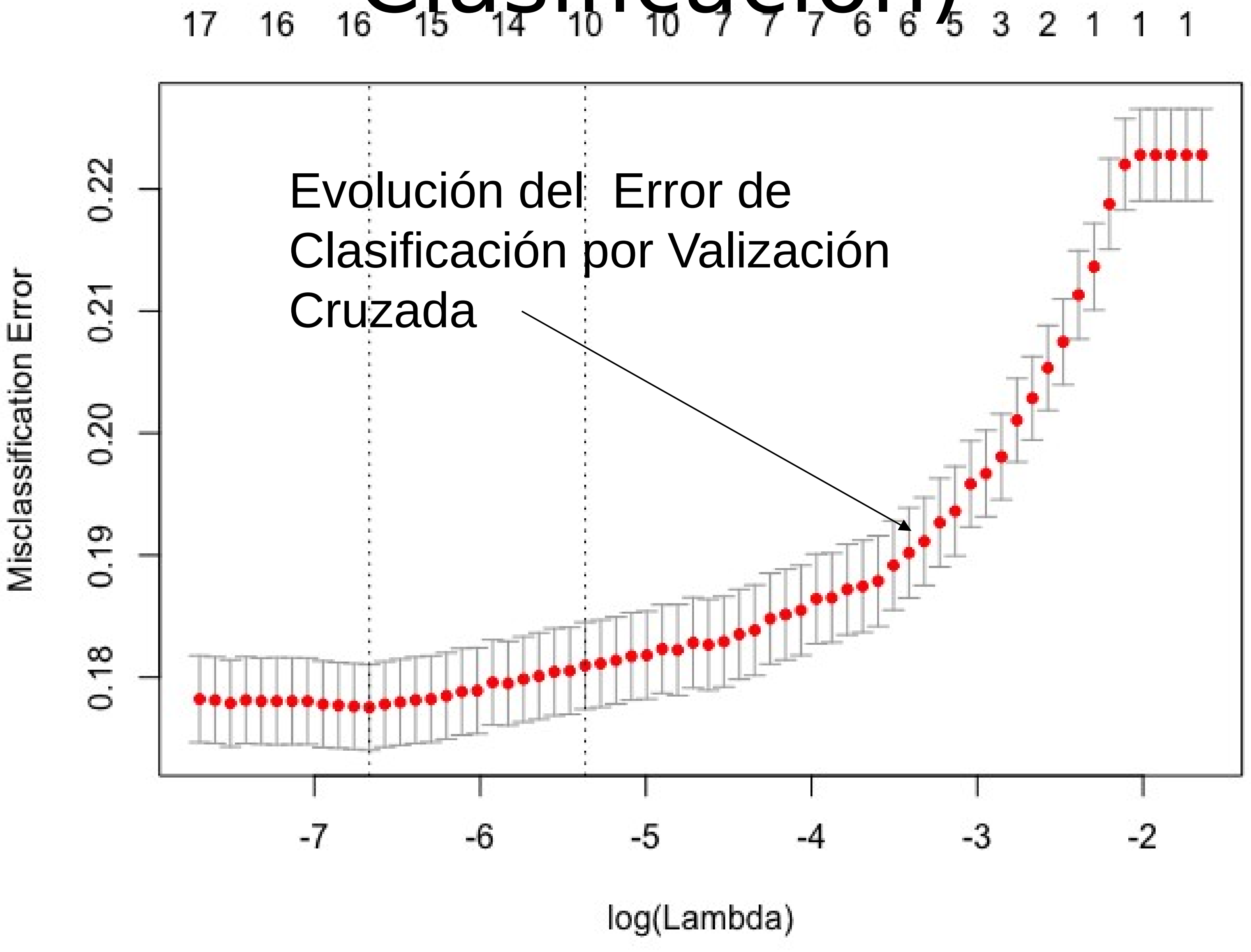
Ejemplo de Regresión Lasso



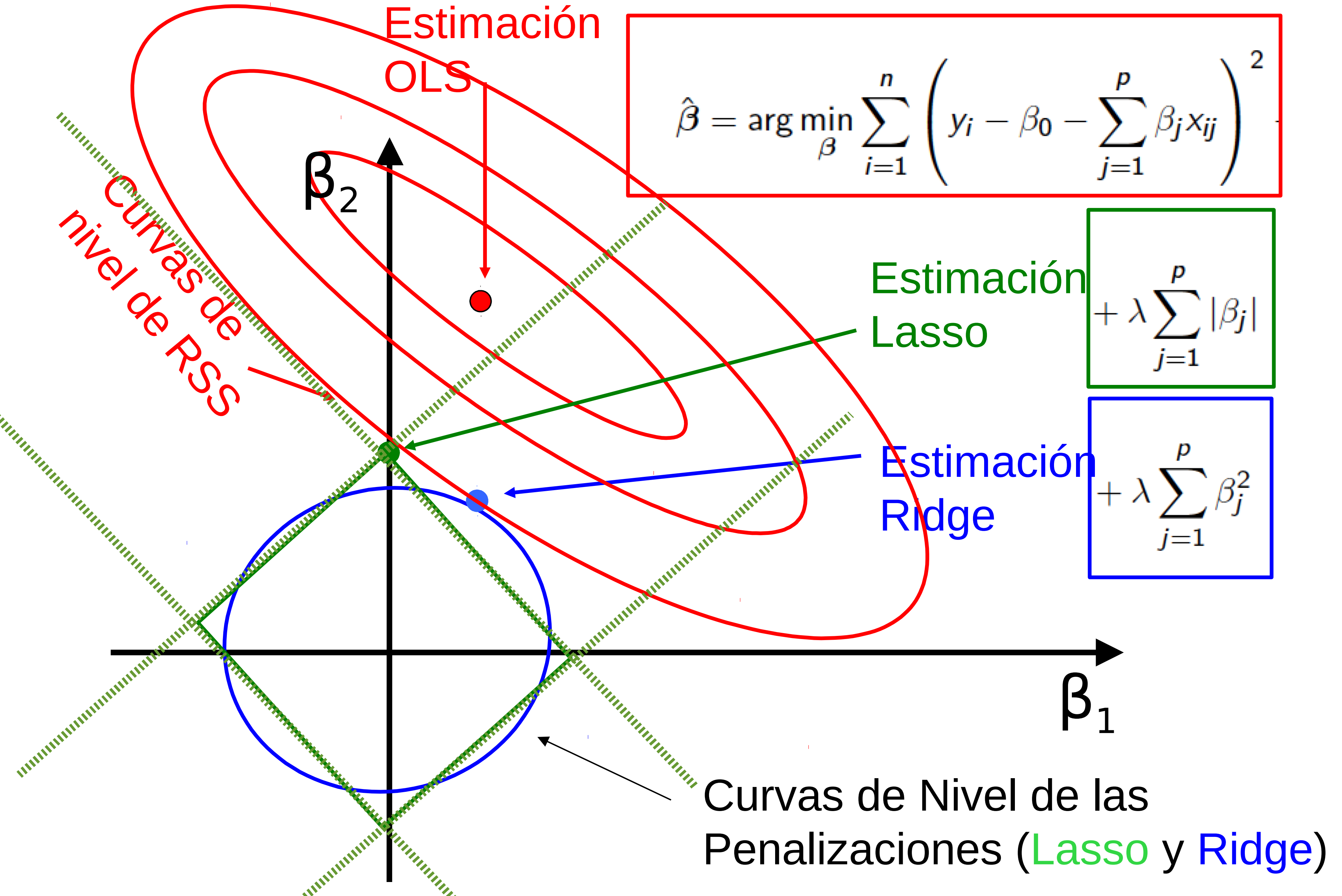
Elección de Lambda (Deviance)



Elección del Lambda (Error Clasificación)



Ridge y Lasso - Intuición Geométrica



El Paquete "glmnet"

Parámetro de Elasticnet

`glmnet` solves the following problem

$$\min_{\beta_0,\beta} \frac{1}{N} \sum_{i=1}^{N} w_i l(y_i, \beta_0 + \beta^T x_i) + \lambda \left[(1-\alpha)||\beta||_2^2/2 + \alpha||\beta||_1\right],$$

over a grid of values of λ covering the entire range. Here $l(y, \eta)$ is the negative log-likelihood contribution for observation i; e.g. for the Gaussian case it is $\frac{1}{2}(y-\eta)^2$. The *elastic-net* penalty is controlled by α, and bridges the gap between lasso ($\alpha = 1$, the default) and ridge ($\alpha = 0$). The tuning parameter λ controls the overall strength of the penalty.