Regularización de Modelos

Los modelos lineales suelen ser más fáciles de interpretar y tiene mayor capacidad de generalización al ser menos flexibles. Es bien sabido que a más variables el modelo es más flexible y por ende su varianza es mayor. Por ello vamos a analizar métodos de regularización de los modelos de ajuste que logren que la cantidad de variables predictoras sea menor, y así obtener una reducción de la varianza del modelo y su consiguiente capacidad de generalización

- **Regularización: Regresión Contraída**

El primer método de regularización o encogimiento (*shrinkage)* es el de Regresión Contraída, o *Ridge Regression* en inglés . La regresión contraída mantiene todas las variables predictores pero disminuye su influencia en la predicción.

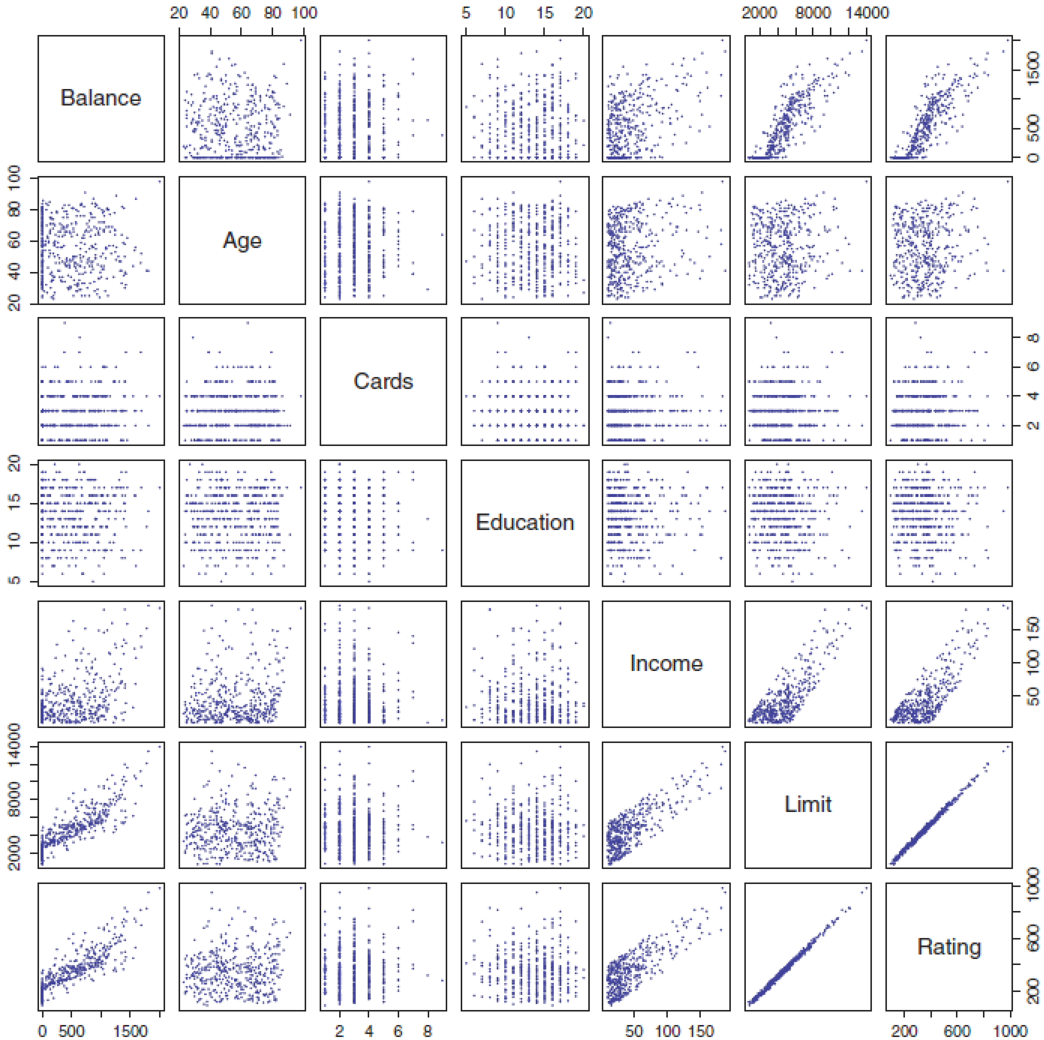
En RL queremos minimizar esta función por Camín $$\mathrm{RSS} = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$$

En RL Contraída le adicionamos :

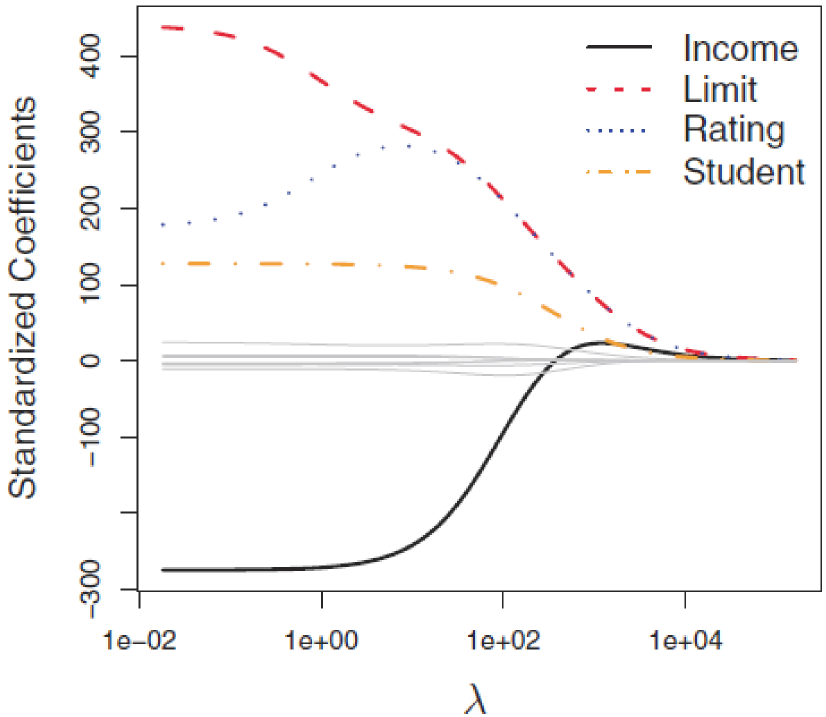
$$\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 = \mathrm{RSS} + \lambda \sum_{j=1}^{p} \beta_j^2$$

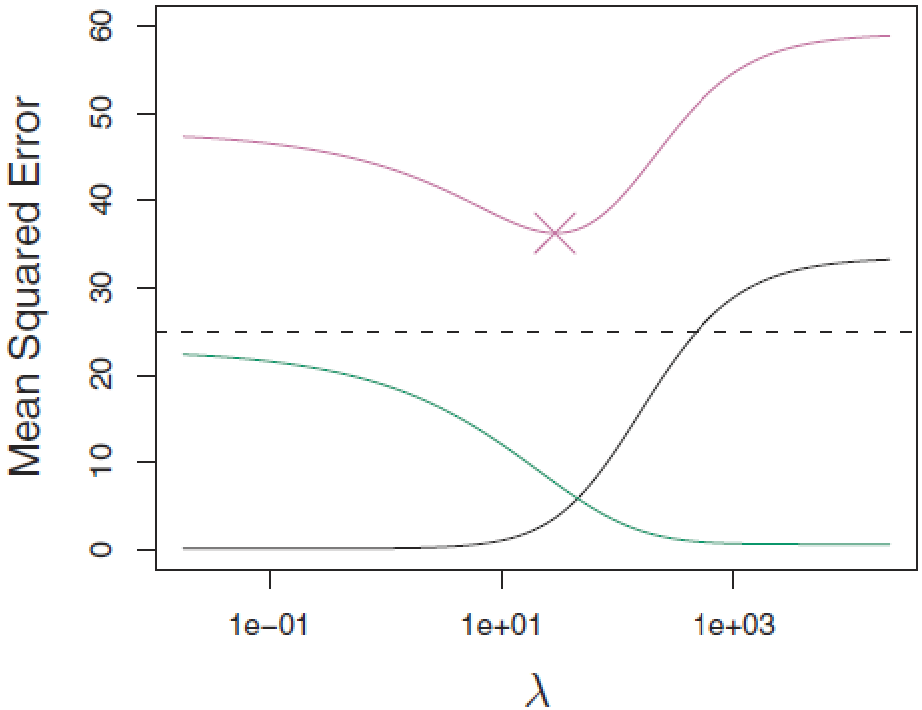
Matriz de Diagramas de dispersión (covarianzas)

Balance ~ age, cards, education , income , limit, rating , vender, student, status , ethnicity



Balance ~ age, cards, education , income , limit, rating , vender, student, status , ethnicity



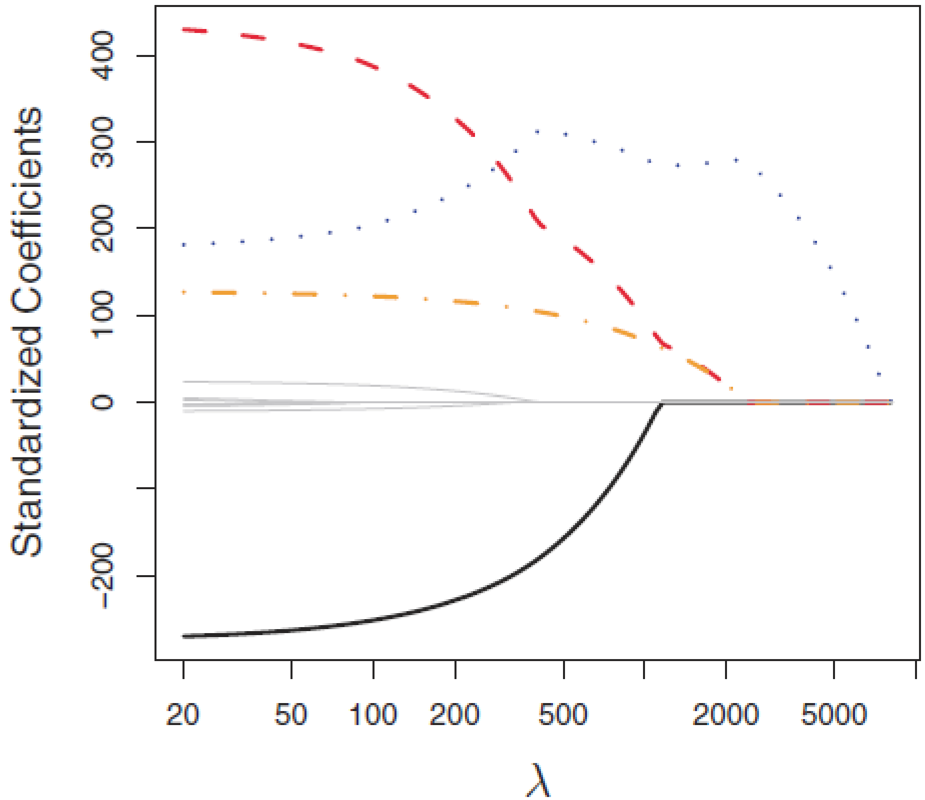


- **Regresión LASSO**

La desventaja de la regresión contraída es que contiene TODOS los predictores. La penalización puede llegar a contraer todos los coeficientes β_j a un valor cercano a cero, pero nunca hacerlos llegar exactamente a cero-.
En LASSO se reemplaza la penalización $|\beta_j|^2$ por $|\beta_j|$

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

Esta nueva penalización si consigue llevar algunos coeficientes a vero, consiguiendo así un modelo con menor varianza y más fácil de interpretar.


Standardized Coefficients
400
300
200
100
0
−200
20
50
100
200
500
2000
5000
λ