

Machine Learning en Economía

Sesión 2: Regularización Ridge y Lasso

Luis Chávez



Lima, 2025

Table of Contents

- 1 Introducción
- 2 Regresión Ridge
- 3 Regresión Lasso
 - Forma unificada
 - Forma generalizada
- 4 Aplicación

Introducción

Introducción

Definición (Análisis de regresión)

Conjunto de técnicas estadísticas y de ML para modelar y analizar la relación entre variables (dependiente e independientes).

Introducción

La función de regresión general típica:

$$E(y_i|x_i) = x_i'\beta \quad (1)$$

El modelo general

$$y_i = x_i'\beta + u_i \quad (2)$$

que asume perturbaciones normales e independiente de las covariables, al estimar por OLS se halla:

$$RSS = \sum_{i=1}^n (y_i - x_i'\beta)^2 \quad (3)$$

¿Problemas en OLS?

- Overfitting¹.
- No permite realizar la selección de predictores.
- No produce coeficientes que sean exactamente cero.
- Interpretaciones laxas del modelo.

¹Gran cantidad de covariables (varianza explosiva) y colinealidad entre ellas.

Introducción

¿Solución? Existen enfoques que permiten efectuar una selección automática de variables: **regularización** o *shrinkage*.

Implicancias:

- Reducción de varianza.
- Incremento de sesgo.
- Interpretación sencilla de coeficientes.

Regresión Ridge

Propuesta por Hoerl y Kennard (1970). La minimización:

$$\min \sum_{i=1}^n (y_i - x_i' \beta)^2 \quad s.a. \quad \sum_{j=1}^k \beta_j^2 \leq s \quad (4)$$

Equivalentemente:

$$\min \sum_{i=1}^n (y_i - x_i' \beta)^2 + \alpha \sum_{j=1}^k \beta_j^2 \quad (5)$$

donde $\alpha \geq 0$ y s son parámetros de ajuste (hiperparámetros).

Generalidades

¿Cómo se elige el parámetro de regularización?

- En R, use *glmnet*.
- En Python use *sklearn*.

Nota (hiperparámetro)

A medida que el parámetro de regularización crece, el sesgo crece pero la varianza decrece.

Regresión Lasso

LASSO (*least absolute shrinkage and selection operator*) fue planteado por Tibshirani (1996).
La minimización:

$$\min \sum_{i=1}^n (y_i - x_i' \beta)^2 \quad s.a. \quad \sum_{j=1}^k |\beta_j| \leq s \quad (6)$$

Equivalentemente,

$$\min \sum_{i=1}^n (y_i - x_i' \beta)^2 + \alpha \sum_{j=1}^k |\beta_j| \quad (7)$$

Unificación

El componente de penalización de ridge se conoce como l_2 y el de lasso se conoce como l_1 .
Una forma de unificación:

$$\min RSS + \lambda \sum_{j=1}^k |\beta_j|^d \quad (8)$$

Se tiene:

- $d=0$ corresponde a un problema de selección de variables.
- $d=1$ corresponde a LASSO.
- $d=2$ corresponde a Ridge.

LASSO generalizado

Las generalizaciones de LASSO son *least angle regression* o LARS (planteado por Efron et al., 2004) y *elastic net* (definido por Zou y Hastie, 2005). Elastic net combina las ventajas de Ridge y LASSO,

$$\min RSS + \alpha \left(\frac{1 - \alpha}{2} \sum_{j=1}^k \beta_j^2 + \alpha \sum_{j=1}^k \beta_j^2 |\beta_j| \right) \quad (9)$$

donde el hiperparámetro $0 \leq \lambda \leq 1$.

Atención

Es importante estandarizar las variables antes de aplicar las herramientas de regularización.

Aplicación

Caso aplicado

Ir al cuaderno de colab.

¡Gracias!