

# Regresión LASSO L1 (Operador de Selección y Contracción Menor Absolutor)

---

La regresión Lasso es una técnica estadística utilizada en el aprendizaje automático para mejorar la precisión de los modelos predictivos, especialmente cuando se trabaja con grandes conjuntos de datos. Su principal objetivo es prevenir el sobreajuste, un problema que ocurre cuando un modelo se ajusta demasiado a los datos de entrenamiento y no funciona bien con datos nuevos. Lasso lo logra mediante la aplicación de una penalización que reduce la magnitud de los coeficientes de las variables predictoras, incluso forzando algunos a cero. Esto simplifica el modelo, selecciona automáticamente las variables más importantes y mejora su capacidad de generalización.

*Aclaración: Aunque la regresión Lasso y la regularización L1 suelen mencionarse como si fueran lo mismo, en realidad no lo son. Lasso es un modelo de regresión que utiliza la penalización L1, pero la regularización L1 también se emplea en otros modelos más allá de la regresión, como en regresión logística, máquinas de soporte vectorial (SVM), redes neuronales y selección de características en modelos lineales. La confusión ocurre porque Lasso es el ejemplo más común de regularización L1, lo que lleva a que muchas fuentes los traten como sinónimos.*

## ¿Qué es la regularización?

En el aprendizaje automático, la regularización es como darle un freno al modelo para que no se vuelva demasiado complicado. El problema es que, a veces, el modelo se aprende de memoria los datos de entrenamiento, como un estudiante que memoriza respuestas exactas para un examen. Esto se llama "sobreajuste". Cuando esto pasa, el modelo funciona mal con datos nuevos. La regularización añade una "penalización" a la fórmula que el modelo usa para saber cuando se equivoca, para que aprenda de manera más general, y no tan específicamente, y de esta manera, no se memorice los datos de entrenamiento.

## ¿Qué es la regresión Lasso?

La regresión Lasso, también conocida como regularización L1, es un tipo de regresión lineal que añade una penalización basada en el valor absoluto de los coeficientes de las variables predictoras. Esta penalización fuerza a algunos coeficientes a ser exactamente cero, lo que significa que esas variables se eliminan del modelo.

## ¿Por qué se utiliza regresión Lasso?

- **Prevención del sobreajuste:** Reduce la complejidad del modelo, evitando que se ajuste demasiado a los datos de entrenamiento.

- Selección automática de variables: Identifica y retiene solo las variables más relevantes para la predicción, simplificando el modelo y mejorando su interpretabilidad.
- Manejo de multicolinealidad: Aunque no la elimina por completo, Lasso puede mitigar los efectos de la multicolinealidad (alta correlación entre variables predictoras) al seleccionar una variable y eliminar las correlacionadas

## ¿Cómo funciona la regresión LASSO?

1. Penalización L1: Lasso añade un término de penalización a la función de coste del modelo de regresión lineal. Este término es proporcional a la suma de los valores absolutos de los coeficientes de las variables predictoras.
2. Parámetro Lambda ( $\lambda$ ): Este parámetro controla la fuerza de la penalización.
  - Un  $\lambda$  alto aumenta la penalización, lo que reduce más coeficientes a cero, simplificando el modelo.
  - Un  $\lambda$  bajo disminuye la penalización, conservando más variables en el modelo.
3. Minimización del Error Cuadrático Medio (MSE): Lasso busca el valor de los coeficientes que minimiza el MSE (la diferencia entre los valores predichos y los reales), sujeto a la penalización L1.
4. Selección de Variables: Al forzar algunos coeficientes a cero, Lasso realiza una selección automática de variables, quedándose solo con las más importantes.

La función de costo que la regresión LASSO trata de minimizar es:

$$J(\beta) = \text{Error} + \lambda \sum_{i=1}^p |\beta_i|$$

Donde:

- $J(\beta)$  es la función de costo.
- **Error** es una medida del error entre los valores predichos y los valores reales (por ejemplo, la suma de los errores al cuadrado en la regresión lineal estándar).
- $\lambda$  (lambda) es el **parámetro de regularización** (un hiperparámetro que se ajusta). Controla la fuerza de la penalización. Cuanto mayor sea el valor de  $\lambda$ , mayor será la penalización.
- $\sum_{i=1}^p |\beta_i|$  es la **norma L1** de los coeficientes, que es la suma de los valores absolutos de todos los coeficientes ( $\beta_i$ ).
- $p$  es el número de variables predictoras.

## Cuándo se utiliza la regresión Lasso?

- Conjuntos de datos de alta dimensión: Cuando hay muchas más variables predictoras que observaciones.
- Selección automática de variables: Cuando se necesita identificar las variables más relevantes para la predicción.

- Problemas predictivos: Donde el objetivo principal es la precisión de la predicción.

## La clave de la regresión LASSO: Selección de características

La principal característica distintiva de la regularización L1 (y por lo tanto de la regresión LASSO) es su capacidad para **forzar algunos de los coeficientes de las variables predictoras a ser exactamente cero**.

- Cuando  $\lambda$  es suficientemente grande, la penalización L1 puede hacer que los coeficientes de las variables menos importantes se reduzcan a cero, eliminando efectivamente esas variables del modelo.
- Esto convierte a la regresión LASSO en un método útil para la **selección de características**, ya que identifica automáticamente las variables más relevantes para la predicción.
- El modelo resultante es más **parsimonioso** (tiene menos variables) y, por lo tanto, a menudo más fácil de interpretar.

## Ventajas de la regresión LASSO:

- Selección de características automática: Identifica y elimina variables irrelevantes, simplificando el modelo y facilitando su interpretación. Por ejemplo, al predecir el precio de una casa, LASSO selecciona las características más relevantes como el tamaño y la ubicación.
- Ayuda a prevenir el sobreajuste: Reduce la complejidad del modelo al penalizar los coeficientes grandes, lo que mejora la precisión en datos nuevos.
- Útil en conjuntos de datos de alta dimensión: Funciona bien cuando hay muchas variables predictoras (alta dimensión), algunas de las cuales pueden ser irrelevantes.
- Mejora la interpretabilidad del modelo: Al tener menos variables, el modelo es más fácil de entender y explicar.
- Puede manejar cierta multicolinealidad: Tiende a seleccionar una variable de un grupo de variables altamente correlacionadas y a establecer los coeficientes de las otras en cero. Por ejemplo, si tienes variables como "metros cuadrados" y "número de habitaciones", LASSO selecciona la más importante.

## Desventajas de la regresión LASSO:

- Puede descartar variables relevantes: Si  $\lambda$  es demasiado grande, podría eliminar variables que realmente tienen un impacto en la predicción, lo que reduciría la precisión del modelo.
- Selección arbitraria en alta multicolinealidad: Si hay grupos de variables muy correlacionadas, LASSO puede seleccionar una de ellas arbitrariamente (al azar), lo que puede ser inestable, ya que pequeñas variaciones en los datos podrían cambiar la variable seleccionada.

- Puede no funcionar tan bien como Ridge si todas las variables son relevantes: Si la mayoría de las variables tienen algún impacto en la predicción, la regresión Ridge podría dar mejores resultados en términos de precisión predictiva.
- La elección del parámetro  $\lambda$  es crucial: Un valor incorrecto de  $\lambda$  puede llevar a un modelo subajustado (si  $\lambda$  es demasiado grande) o sobreajustado (si  $\lambda$  es demasiado pequeño). La selección óptima de  $\lambda$  a menudo se realiza mediante técnicas de validación cruzada.

## Casos de Uso Comunes

- Genómica: Identificación de genes relevantes para ciertas enfermedades.
- Marketing: Selección de las variables demográficas y de comportamiento más importantes para predecir la respuesta a una campaña publicitaria.
- Finanzas: Construcción de modelos de riesgo crediticio, seleccionando los factores más influyentes.

## Diccionario de Términos Clave

- **Sobreajuste (Overfitting):** Un modelo que se ajusta demasiado a los datos de entrenamiento y no generaliza bien a datos nuevos.
- **Regularización:** Técnicas para prevenir el sobreajuste, añadiendo información adicional para evitar la complejidad excesiva del modelo.
- **Penalización L1:** La penalización utilizada en Lasso, basada en el valor absoluto de los coeficientes.
- **Lambda ( $\lambda$ ):** El parámetro que controla la fuerza de la penalización en Lasso.
- **Alfa ( $\alpha$ ):** Este término no se menciona en el texto proporcionado.
- **Multicolinealidad:** Alta correlación entre variables predictoras.
- **Error Cuadrático Medio (MSE):** Una medida de la diferencia entre los valores predichos y los reales.
- **Sesgo:** La diferencia entre las predicciones promedio de un modelo y los valores reales.
- **Varianza:** La variabilidad de las predicciones de un modelo para diferentes conjuntos de datos.
- **Función de Coste:** Una función que mide qué tan bien un modelo se ajusta a los datos; el objetivo es minimizar esta función.
- **Coefficientes Beta ( $\beta$ ):** Los valores que multiplican las variables predictoras en un modelo de regresión.
- **R2:** Una medida de qué proporción de la varianza en la variable dependiente es explicada por las variables independientes.
- **Variable Dependiente:** La variable que se está prediciendo.
- **Variables Independientes:** Las variables utilizadas para hacer predicciones.
- **Elastic Net:** Una técnica de regularización que combina las penalizaciones L1 (Lasso) y L2 (Ridge).
- **Ridge:** También conocida como regularización L2.

---

---

## Ejemplo: ☕

Tenemos una cafetería y queremos saber qué cosas hacen que venda más café por día.

Hemos estado anotando datos durante varias semanas, y tenemos variables como:

- ☒ Hace frío
- ☒ Es fin de semana
- ☒ Hay promociones
- ☒ Hay música en vivo
- ☒ Se encendieron luces decorativas
- ☒ Cambio en el menú
- ☒ Hay sillas nuevas
- ☒ Hubo luna llena 🌕
- ☒ El gato se subió al mostrador 🐱

---

## 🎯 Nuestro objetivo:

**Predecir cuántos cafés se venderán**, dependiendo de nuestros factores factores.

Pero hay un problema:

👉 No todos esos factores realmente influyen. Algunos son importantes (como promociones), otros son ruido (como la luna llena o el gato 🐱).

---

## 📊 Aquí entra la regresión Lasso (L1)

Lasso es un tipo de **regresión lineal con regularización**. Es como un filtro que te ayuda a:

- 🧠 Seleccionar solo las variables que realmente ayudan a predecir
- ✂ Eliminar o ignorar las que no sirven
- 📄 Hacer el modelo más simple y más generalizable

---

## 📊 ¿Cómo funciona en la cafetería?

Digamos que alimentamos al modelo con registros de muchos días:

- Día 1: Hace frío + fin de semana + promoción + música = 280 cafés
- Día 2: Calor + lunes + sin promo + gato en el mostrador = 90 cafés

- Día 3: Frío + cambio de menú + luces + música = 200 cafés
- Día 4: Luna llena + sillas nuevas + promo = 100 cafés

El modelo va **aprendiendo qué cosas realmente impactan en las ventas**.



## Modelo clásico vs. Lasso



## Modelo de regresión lineal (sin penalización):

$$\hat{y} = \beta_0 + \beta_1 \cdot \text{Frío} + \beta_2 \cdot \text{FinDeSemana} + \beta_3 \cdot \text{Promoción} + \beta_4 \cdot \text{Música} + \beta_5 \cdot \text{Luces} + \beta_6 \cdot \text{Luna Llena}$$

$\hat{y}$  corresponde a las ventas predichas.

Este modelo podría dar pesos ( $\beta_i$ ) pequeños pero no exactamente cero a variables irrelevantes como la luna llena o el gato.



## Regresión Lasso (con penalización L1):

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

que sería:

$$\text{Coste} = \underbrace{\frac{1}{2n} \sum_{i=1}^n (\text{Ventas}_i - \text{VentasPredichas}_i)^2}_{\text{Qué tan lejos están nuestras predicciones de las ventas reales}} + \underbrace{\lambda \sum_{j=1}^p |\beta_j|}_{\text{Castigo por dar importancia a demasiados fact}}$$

$\min_{\beta}$  = El valor que queremos minimizar

$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2$  = **SSE** (suma de errores cuadráticos) Suma los términos que le siguen para cada observación  $i$ , desde la primera ( $i=1$ ) hasta la última ( $i=n$ ).

- $y_i$  = El número real de cafés vendidos en el día  $i$
- $\beta_0$  = El número que el modelo predice para el día  $i$
- $\sum_{j=1}^p x_{ij} \beta_j$  = contribución de todas las variables donde para una observación específica  $i$ , multiplica el valor de cada variable predictora  $j$   $x_{ij}$  por su correspondiente coeficiente  $\beta_j$  y luego suma todos estos productos para todas las  $p$  variables predictoras.
  - $\sum$  = sumatoria
  - $\sum_{j=1}^p$  = Son los límites de la sumatoria
    - $j$  = índice de la sumatoria

- $=1$  = indica que la sumatoria comienza con el valor de  $j$  igual a 1 (primera variable predictora del conjunto de datos)
- $p$  = indica que la sumatoria termina cuando el valor de  $j$  alcanza  $p$ . En donde  $p$  representa el número total de variables predictoras en el modelo
- $x_{ij}$  = representa el valor de la  $j$ -ésima variable predictora para la  $i$ -ésima observación
- $\beta_j$  = beta con subíndice  $j$  representa el coeficiente asociado a la  $j$ -ésima variable predictora.
- $^2$  = el error es elevado al cuadrado para: 1.- evitar la cancelación de errores positivos y negativos. 2.- Penaliza más los errores grandes. 3.- Facilita el cálculo

$\frac{1}{2n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \text{MSE}$  (error cuadrático medio). Cómo la SSE pero añadiendo el factor escala que mide qué tan lejos están nuestras predicciones de los valores reales

- $y_i$  = Valor real de la variable dependiente (números de cafés vendidos)
- $\hat{y}_i$  = Valor predicho de la variable dependiente por el modelo de regresión para la  $i$ -ésima observación (números de cafés que el modelo predice que se venderán el día  $i$ )
- $^2$  = El error es elevado al cuadrado para: 1.- evitar la cancelación de errores positivos y negativos. 2.- Penaliza más los errores grandes. 3.- Facilita el cálculo

$\lambda \sum_{j=1}^p |\beta_j|$  = es la **penalización L1**. Esta penalización fuerza a los coeficientes de las variables menos importantes a ser exactamente **cero**.

- $\lambda$  = Parámetro de regularización o constante de penalización
  - Si es cercano a 0, la penalización es débil (sin mucha tendencia a reducir los coeficientes a cero)
  - Si es grande, la penalización es fuerte (modelo incentivado a hacer muchos de los coeficientes de las variables predictoras cero)
  - Valor óptimo : Se encuentra generalmente mediante técnicas de validación cruzada.
- $\sum$  = sumatoria
  - $\sum_{j=1}^p$  = Son los límites de la sumatoria
    - $j$  = índice de la sumatoria
    - $=1$  = indica que la sumatoria comienza con el valor de  $j$  igual a 1 (primera variable predictora del conjunto de datos)
    - $p$  = indica que la sumatoria termina cuando el valor de  $j$  alcanza  $p$ . En donde  $p$  representa el número total de variables predictoras en el modelo
- $\beta_j$  = Coeficiente (el peso) asociado con la  $j$ -ésima variable predictora del modelo. Indica la fuerza y dirección de la relación entre las variables predictora y dependiente (impacto considerable o no en la venta de café).
- $|\cdot|$  = Símbolo de valor absoluto, es decir, se toma el valor (número) absoluto independientemente de si es negativo o positivo ( $-3=3$ ,  $3=3$ )

Después de entrenar el modelo Lasso con los datos, podríamos obtener algo como:

$$\text{Ventas} \approx 50 + 80 \cdot \text{Frío} + 120 \cdot \text{FinDeSemana} + 150 \cdot \text{Promoción} + 60 \cdot \text{Música} + 0 \cdot \text{Lu}$$

Y esto nos dice claramente que:

- 🚀 El **frío**, los **fines de semana** y las **promociones** son factores clave que aumentan tus ventas.
  - 🎵 La **música en vivo** y un **nuevo menú** tienen un impacto menor pero positivo.
  - 💡 Las **luces decorativas**, las **sillas nuevas**, la **luna llena** y las travesuras de tu **gato** parecen no tener una influencia significativa en las ventas.
- 

### 🚀 En resumen, Lasso nos ayuda a:

- **Identificar los factores de venta reales.**
- **Simplificar tu modelo de predicción.**
- **Tomar decisiones más inteligentes para tu cafetería.**

con esto, ahora sí nos podemos enfocar en los factores que realmente influyen en nuestro negocio haciéndolo crecer! 🍵

---

## Links útiles

<https://www.ibm.com/es-es/think/topics/lasso-regression>