



Pontificia Universidad
JAVERIANA
Bogotá

[VIGILADA MINEDUCACIÓN]

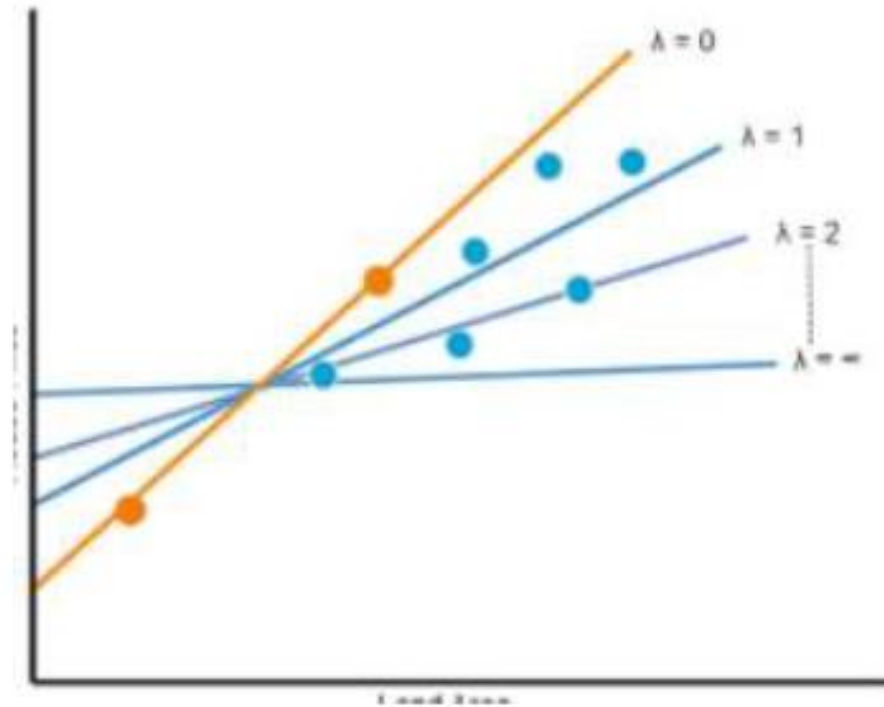
Regularización en Regresión Lineal



Pontificia Universidad
JAVERIANA
Colombia

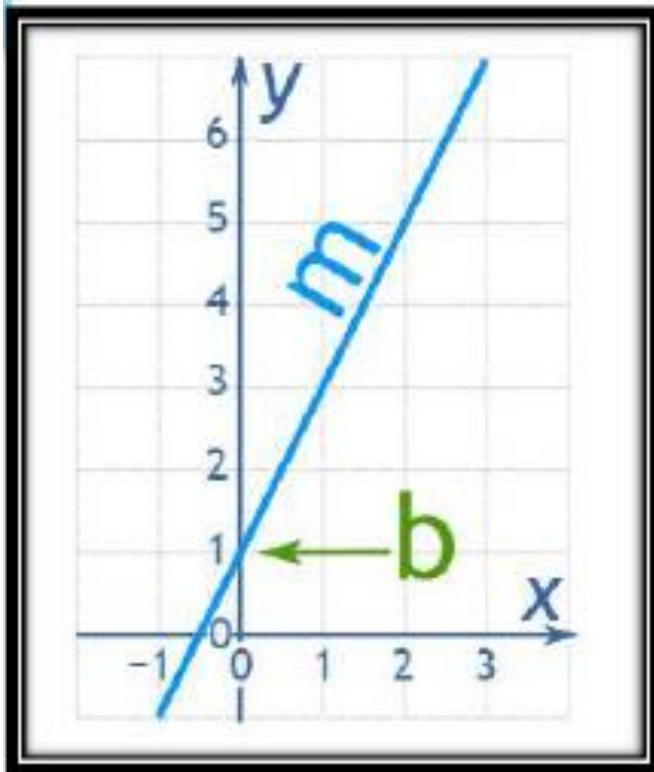
Técnicas de Regularización en Regresión Lineal

Lasso, Ridge & Elastic Net Regression



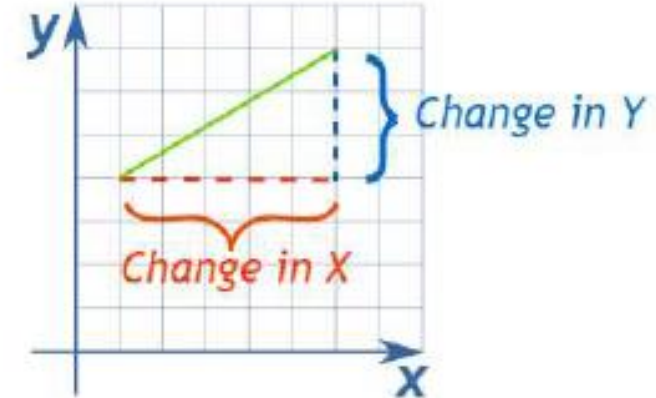
Ecuación de la línea recta

$$y = mx + b$$



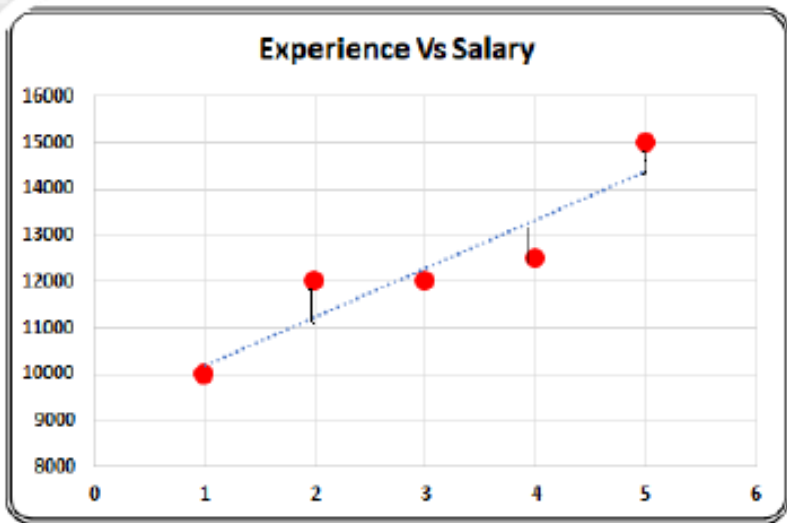
- b = valor de y cuando $x = 0$ (intercepto)
- m = pendiente o gradiente (qué tan inclinada está la línea)

$$m = \frac{\text{Change in Y}}{\text{Change in X}}$$



Ecuación de la línea recta

La regresión lineal es un método para modelar la relación entre una variable dependiente y una o más variables independientes.



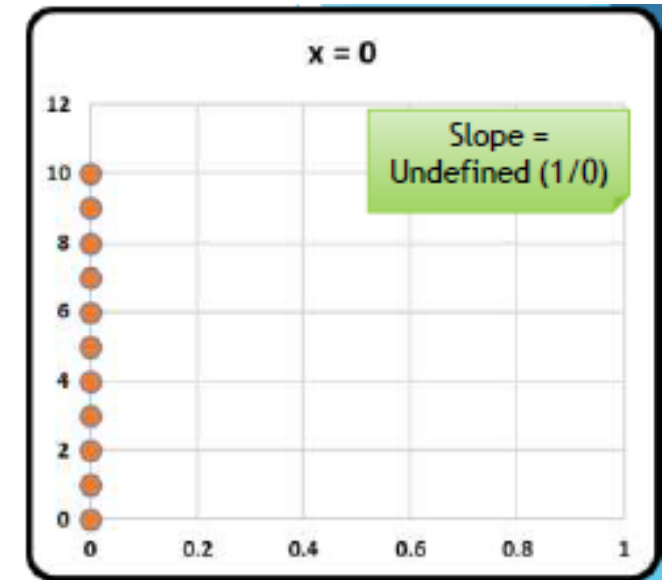
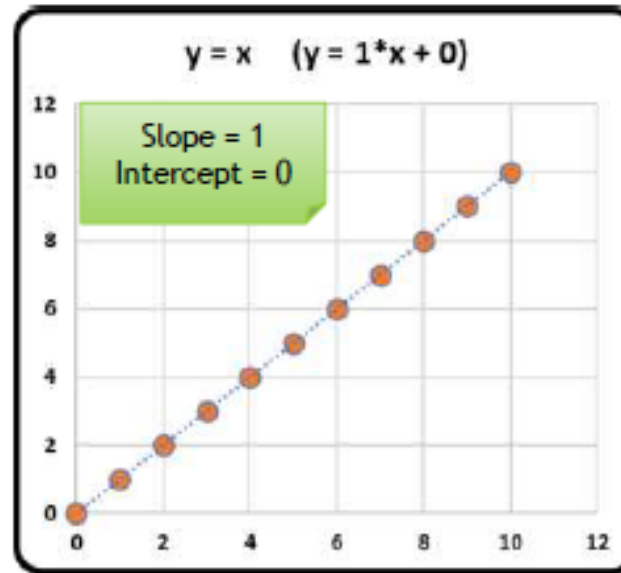
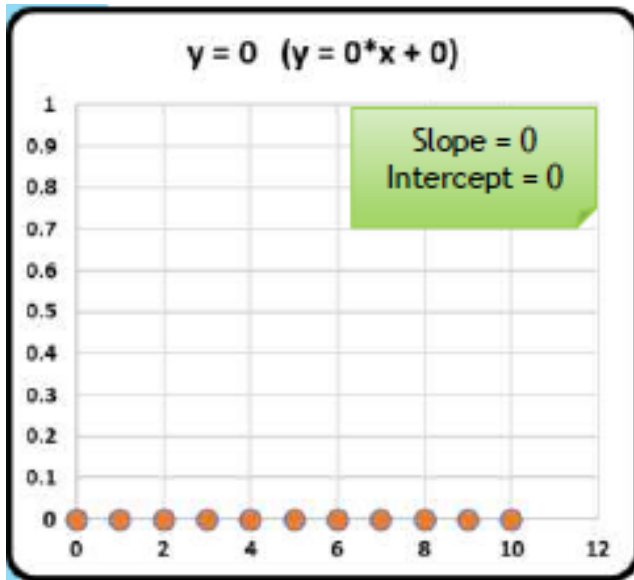
- El modelo de regresión lineal intenta crear una relación lineal entre las variables dependiente (salario) e independiente (experiencia).
- Intenta crear la ecuación de una línea recta ($\text{salario} = m * \text{experiencia} + b$) con un error mínimo (residual) entre el valor real y el valor predicho.

La regresión lineal intenta crear la línea de mejor ajuste con la mínima suma de residuos ($\sum (Y - Y_{\text{predict}})^2$) que también conocida como función de costo o pérdida

Ecuación de regresión lineal simple: $y = m * X + b$

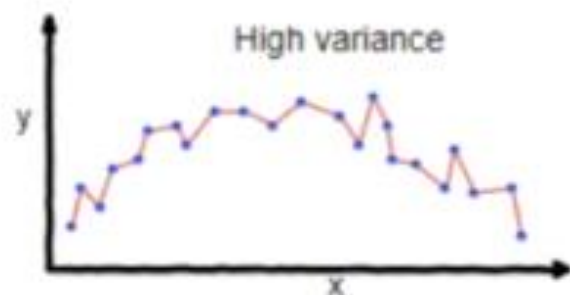
Ecuación de regresión lineal múltiple: $y = \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_n * X_n + \beta_0$

Impacto de la pendiente

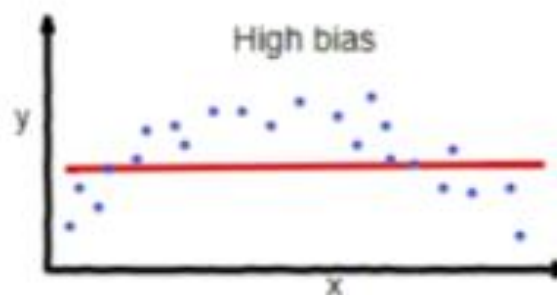


- En el caso de las rectas de mayor pendiente, cualquier variación menor en x puede causar una variación en y .
- En el modelo de regresión lineal que se crea con tales datos puede haber un problema de sobreajuste.
- Una solución es penalizar las pendientes y generalizar el modelo.
- La regresión LASSO y RIDGE son modelos que ayudan para lo mismo.

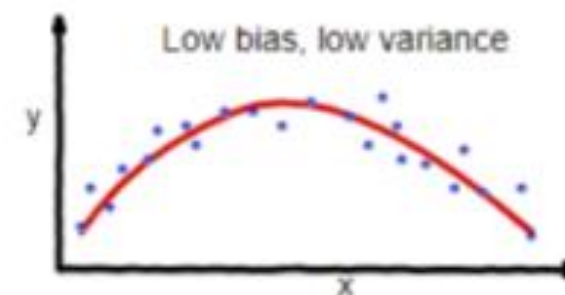
Impacto de la pendiente



overfitting



underfitting



Good balance

UNDERFITTING se produce cuando un modelo no puede capturar con precisión las dependencias entre los datos, normalmente como consecuencia de su propia simplicidad. Suele tener buen desempeño con datos conocidos y una mala generalización con datos nuevos.

OVERFITTING se produce cuando un modelo aprende demasiado bien las dependencias entre los datos. Usualmente tienen un alto R^2 . Sin embargo, suelen generalizar bien con datos conocidos y tienen un R^2 significativamente menor con datos nuevos.

Modelo Lineal Regularizado



- Una buena forma de reducir el OVERFITTING es regularizar el modelo (es decir, restringirlo):
- Para un modelo lineal, la regularización se consigue normalmente restringiendo los pesos del modelo. ¿Por qué?
- El tamaño de los coeficientes aumenta exponencialmente al aumentar la complejidad del modelo.
- Técnicas como *Ridge Regression*, *Lasso Regression* y *Elastic Net*, implementan tres formas diferentes de restringir los pesos.

Regresión LASSO

Least Absolute Shrinkage and Selection Operator

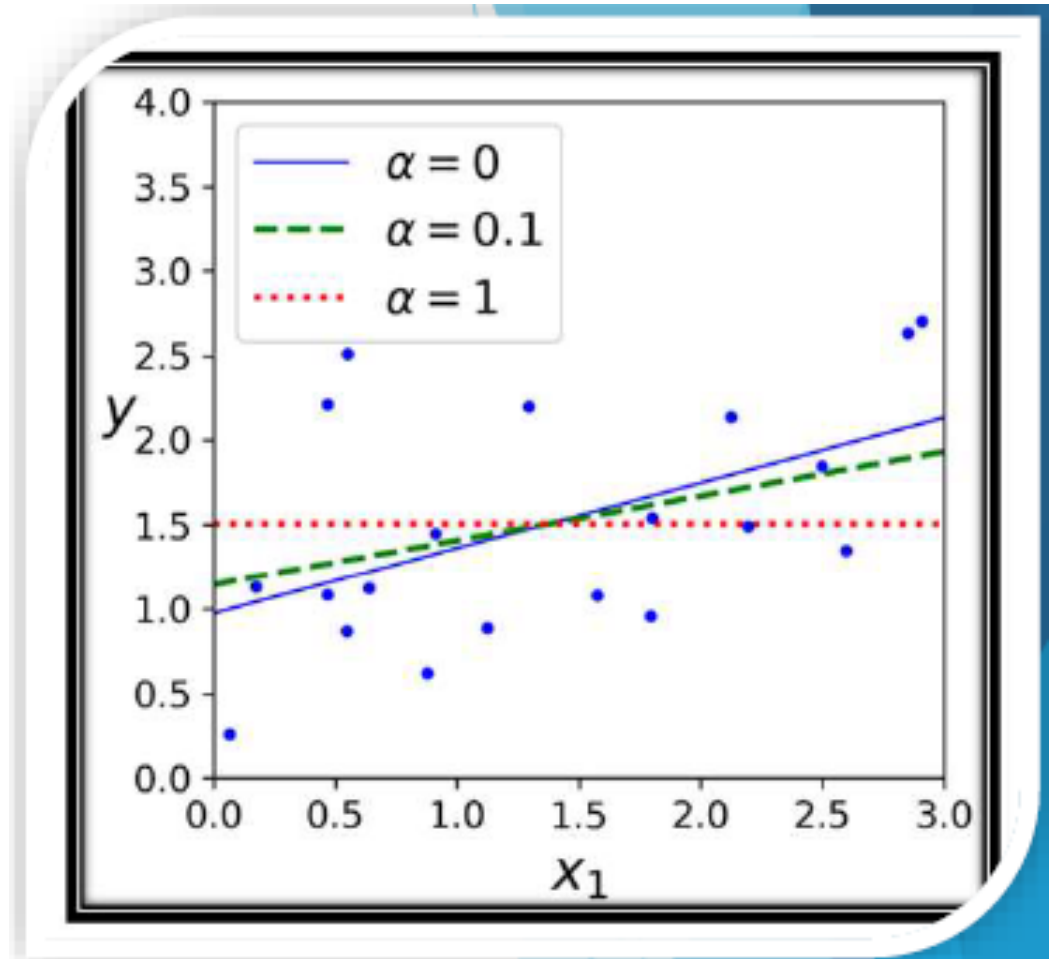
Regression (Lasso) es otra versión regularizada de regresión lineal: añade un término de regularización L1 a la función de costo,

Los objetivos de minimización son $MSE + \alpha * |m|$

El valor por defecto de α es 1. Puede ser cero o cualquier número positivo.

La regresión Lasso tiende a eliminar completamente los pesos de las variables menos importantes (los lleva a cero).

- α denota la cantidad de encogimiento (*shrinkage*)
- $\alpha = 0$ implica que todas las variables son consideradas y es equivalente al modelo de regresión lineal donde solo se considera la suma de cuadrados del error para construir el modelo predictivo
- $\alpha = \infty$ implica que ninguna variable es considerada. A medida que α se aproxima a infinito, elimina más y más variables



Regresión RIDGE

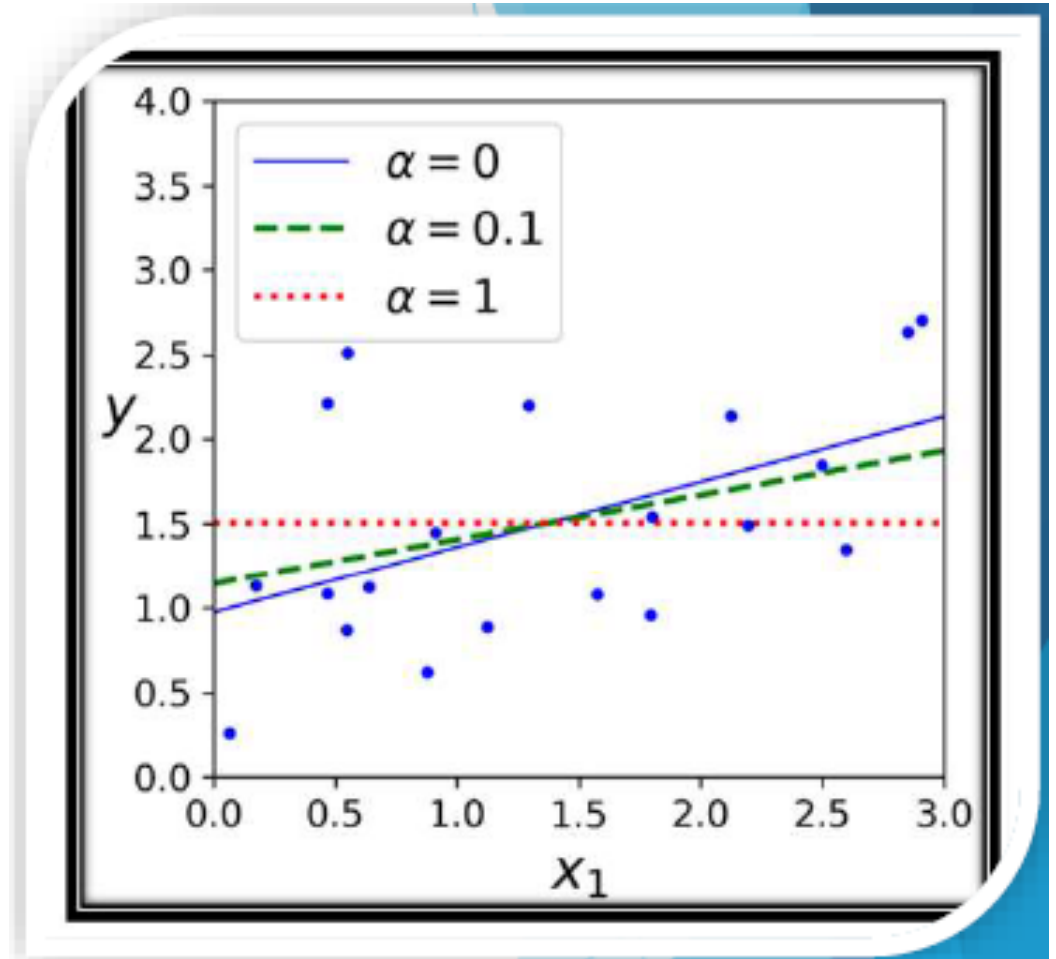
La regresión Ridge (también llamada regularización Tikhonov) es una versión regularizada L2 de la regresión lineal: se añade a la función de costos un término de regularización igual a $\alpha * \text{pendiente}^2$ a la función de costo.

El objetivo de minimización es $\text{MSE} + \alpha * m^2$

Esto obliga al algoritmo de aprendizaje no sólo a ajustar los datos, sino también a mantener los pesos del modelo lo más pequeños posible.

α equilibra la cantidad de énfasis a la regularización. El valor por defecto de α es 1. Puede ser cero o cualquier número positivo.

- $\alpha = 0$ regresión lineal
- $\alpha = \infty$ todos los coeficientes se hacen cero, aumentar α conduce a predicciones más planas (es decir, menos extremas, más razonables).



- La principal diferencia práctica entre *lasso* y *ridge* es que el primero consigue que algunos coeficientes sean exactamente cero, por lo que realiza selección de predictores, mientras que el segundo no llega a excluir ninguno. Esto supone una ventaja notable de *lasso* en escenarios donde no todos los predictores son importantes para el modelo y se desea que los menos influyentes queden excluidos.
- Por otro lado, cuando existen predictores altamente correlacionados (linealmente), *ridge* reduce la influencia de todos ellos a la vez y de forma proporcional, mientras que *lasso* tiende a seleccionar uno de ellos, dándole todo el peso y excluyendo al resto. En presencia de correlaciones, esta selección varía mucho con pequeñas perturbaciones (cambios en los datos de entrenamiento), por lo que, las soluciones de *lasso*, son muy inestables si los predictores están altamente correlacionados.
- Para conseguir un equilibrio óptimo entre estas dos propiedades, se puede emplear la penalización *elastic net*, que combina ambas estrategias.

Regresión de Red Elástica (Elastic Net)

- La red elástica es un tipo popular de regresión lineal regularizada que combina dos penalizaciones populares, concretamente las funciones de penalización L1 (Lasso) y L2 (Ridge).
- Similar a Ridge y LASSO, Elastic Net es una extensión de la regresión lineal, que añade penalizaciones de regularización a la función de pérdida durante el entrenamiento.
- La penalización es una mezcla de las penalizaciones L1 y L2.
- El objetivo de minimización es $MSE + \alpha * |m| + (1 - \alpha) * m^2$

GLMNET

El objetivo de minimización es $MSE + \lambda \{\alpha * |m| + (1 - \alpha) * m^2\}$