

# INTRODUCCIÓN AL APRENDIZAJE AUTOMÁTICO

## TÉCNICAS DE REGULARIZACIÓN – L1 Y L2

LAURA DIAZ DÁVILA – FRANCISCO TAMARIT

# SELECCIÓN DE VARIABLES



BONDAD DE AJUSTE NO ES LO MISMO QUE GENERALIZAR.  
AHORA QUEREMOS PREDECIR, ENFRENTAR EL MODELO A NUEVAS SITUACIONES.



SIEMPRE ES NECESARIO SEPARAR EL DATASET EN TRAINING Y TESTING



LASSO PERMITE ANULAR VARIABLES, POR LO TANTO, SELECCIONAR VARIABLES RELEVANTES: MEJORA LA INTERPRETABILIDAD DE LAS REGRESIONES. TIENE UNA RESOLUCIÓN MÁS COMPLICADA QUE RIDGE YA QUE NO ES DERIVABLE EL RESIDUO LASSO



RIDGE LAS MINIMIZA PERO NUNCA LAS ANULA

# OBJETIVO: MAYOR PODER PREDICTIVO



EVITAR OVERFITTING



REDUCIR VARIANZA (RUIDO)



ATENUAR EFECTO DE CORRELACIÓN ENTRE LAS VARIABLES  
INDEPENDIENTES



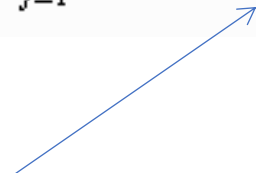
MINIMIZAR LA INFLUENCIA DE LOS PREDICTORES MENOS  
RELEVANTES

# TÉCNICAS DE REGULARIZACIÓN: RIDGE PARA REGRESIÓN LINEAL

ANTES DE COMENZAR ES NECESARIO  
ESCALAR LAS VARIABLES  
INDEPENDIENTES O FEATURES

$$(\|\beta\|_2^2 = \sum_{j=1}^p \beta_j^2)$$

PENALIZA LA SUMA CUADRÁTICA DE LOS  
PARÁMETROS: REDUCE LOS  
COEFICIENTES SIN QUE LLEGUEN A CERO

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


A MEDIDA QUE AUMENTA  $\lambda$ ,  
MAYOR ES LA PENALIZACIÓN Y  
MENOR EL VALOR DE LOS  
PREDICTORES

REDUCE LA VARIANZA FRENTE A OLS. ES  
CAPAZ DE REDUCIR LA VARIANZA SIN  
AUMENTAR EL BIAS, CON LO CUAL  
REDUCE EFECTIVAMENTE EL ERROR  
TOTAL, PARA UN VALOR ADECUADO DE  $\lambda$

DESVENTAJA: EL MODELO FINAL INCLUYE A TODOS LOS  
PREDICTORES.  
PROBLEMAS DE INTERPRETACIÓN CON LOS  
PREDICTORES MENOS INFLUYENTES. NUNCA LOS  
ELIMINA

# TÉCNICAS DE REGULARIZACIÓN: LASSO (TIBSHIRANI- 1996) PARA REGRESIÓN LINEAL

**PENALIZA LA SUMA DE LOS  
PARÁMETROS EN VALOR ABSOLUTO**

$$(\|\beta\|_1 = \sum_{j=1}^p |\beta_j|)$$

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

**A MEDIDA QUE AUMENTA  $\lambda$ , MAYOR ES LA PENALIZACIÓN Y  
MÁS PREDICTORES QUEDAN EXCLUIDOS.**

**FUERZA A LOS PARÁMETROS A QUE TIENDAN A CERO.**

**CONSIGUE EXCLUIR A LAS VARIABLES INDEPENDIENTES  
MENOS RELEVANTES**

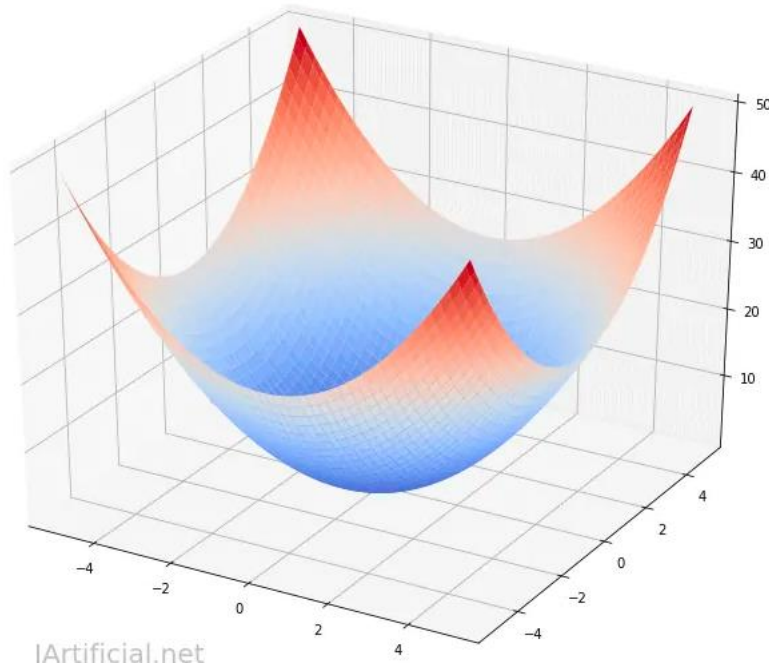
**DESVENTAJA: ES MUY INESTABLE SI LAS VARIABLES  
INDEPENDIENTES ESTÁN ALTAMENTE CORRELACIONADAS**

# REGULARIZACIÓN EN EL DESCENSO POR EL GRADIENTE

$$J = MSE$$

$$J = MSE + \alpha \cdot C$$

**$\alpha$  : SOLUCIÓN DE COMPROMISO  
ENTRE LA SIMPLICIDAD DEL MODELO Y  
SU PERFORMANCE**



**MINIMIZAMOS LA COMPLEJIDAD DEL MODELO Y LA FUNCIÓN DE COSTE**

# LASSO EN EL DESCENSO POR EL GRADIENTE

$$J = MSE$$

$$J = MSE + \alpha \cdot C$$

$\alpha$  : SOLUCIÓN DE COMPROMISO  
ENTRE LA SIMPLICIDAD DEL  
MODELO Y SU PERFORMANCE

SE APLICA EN: REGRESIÓN LOGÍSTICA, POLINÓMICA, REDES  
NEURONALES, MÁQUINAS DE SOPORTE VECTORIAL (SVM)

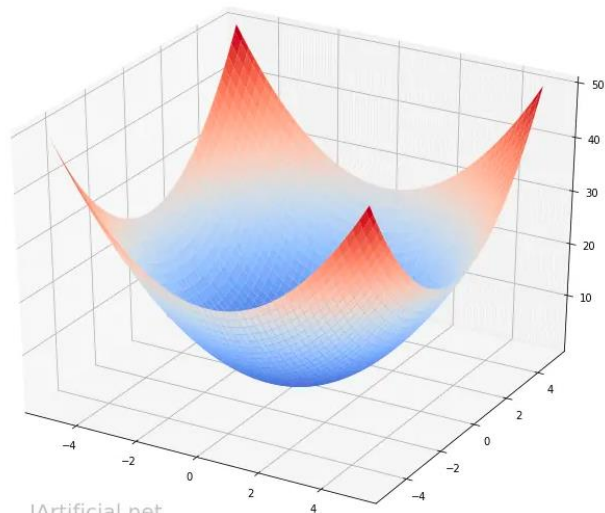
$$J = \frac{1}{M} \sum_{i=1}^M (real_i - estimado_i)^2 + \alpha \frac{1}{N} \sum_{j=1}^N |w_j|$$

¿CUÁNDO USAMOS LASSO?

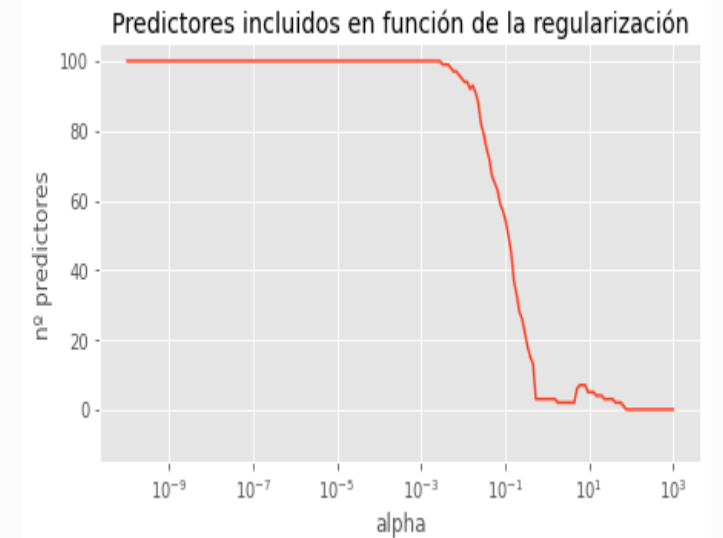
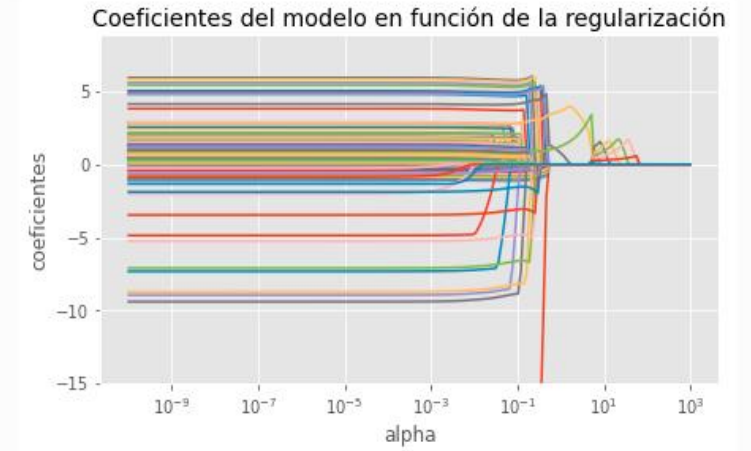
SOSPECHAMOS QUE HAY CARACTERÍSTICAS  
POCO RELEVANTES

QUEREMOS SELECCIONAR FEATURES MÁS  
RELEVANTES

CUANDO LOS ATRIBUTOS NO ESTÁN MUY  
CORRELACIONADOS



IArtificial.net



# RIDGE EN EL DESCENSO POR EL GRADIENTE

$$J = MSE$$

$$J = MSE + \alpha \cdot C$$

SE APLICA EN CUALQUIER MODELO DE MACHINE LEARNING QUE USE  
DESCENSO POR EL GRADIENTE

AHORA C SE EXPRESA COMO:

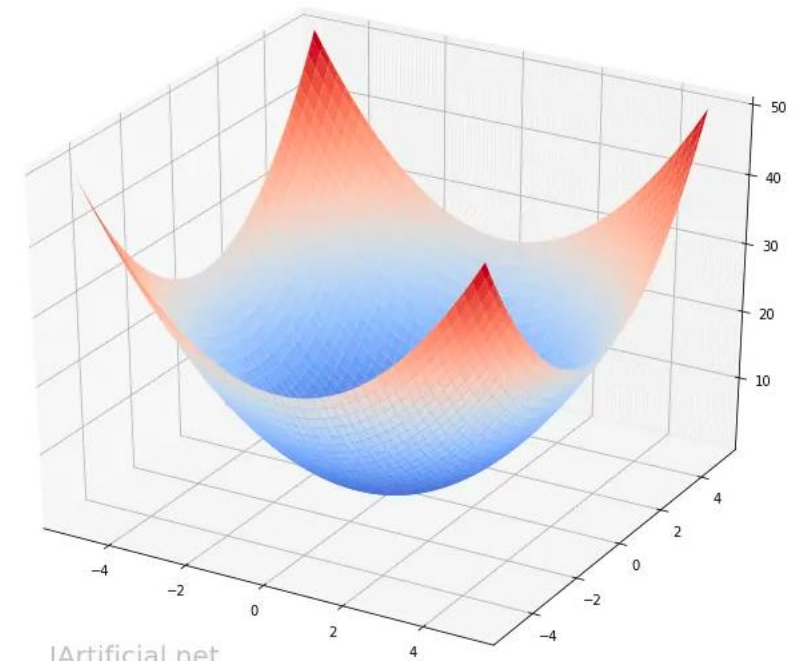
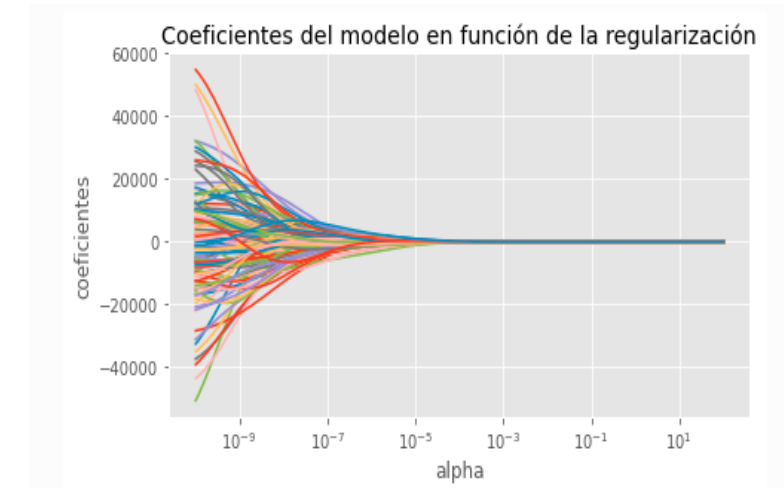
$$C = \frac{1}{2N} \sum_{j=1}^N w_j^2$$

$$J = \frac{1}{M} \sum_{i=1}^M (real_i - estimado_i)^2 + \alpha \frac{1}{2N} \sum_{j=1}^N w_j^2$$

¿CUÁNDO USAMOS RIDGE? :

CUANDO TENEMOS SOSPECHAS  
QUE LOS ATRIBUTOS ESTÁN  
CORRELACIONADOS

CUANDO LA MAYORÍA DE LOS  
ATRIBUTOS (FEATURES) SON  
RELEVANTES

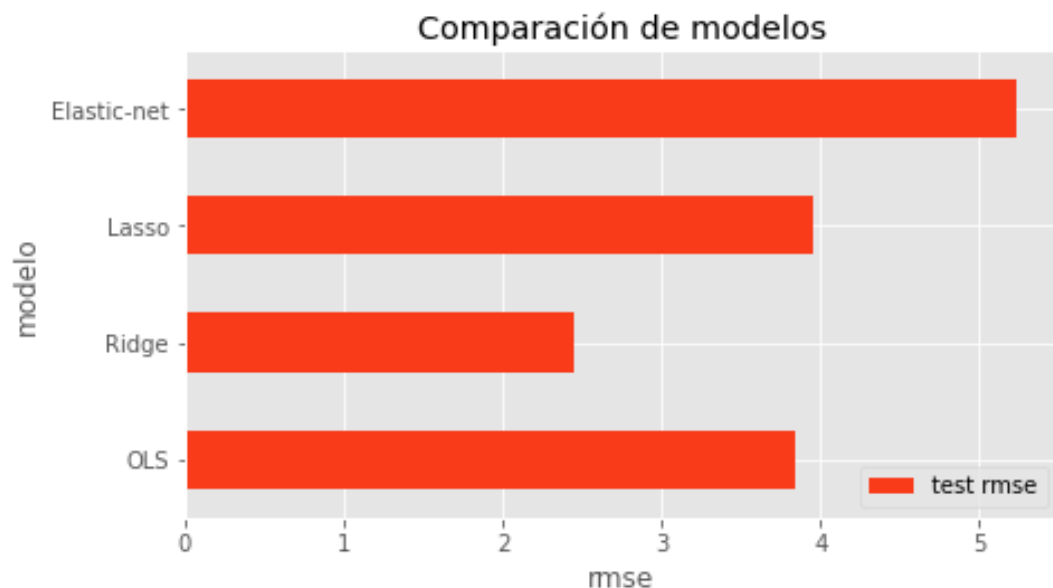




# UNA MEJORA: ELASTIC NET

**INTRODUCE OTRO HIPERPARÁMETRO:  $\alpha$ .**  
**CUANDO VALE CERO APLICA RIDGE,**  
**CUANDO VALE UNO, LASSO**

$$(\alpha\lambda||\beta||_1 + \frac{1}{2}(1 - \alpha)||\beta||_2^2)$$



**SE MATERIALIZA EN UNA COMBINACIÓN DE  
AMBAS TÉCNICAS DE REGULARIZACIÓN, QUE  
SUELE DAR MUY BUENOS RESULTADOS**

# INVITACIÓN A PROFUNDIZAR

**LAS LIBRERÍAS MÁS UTILIZADAS PARA REGRESIÓN LINEAL:**

Scikit-learn  
[statsmodels](#)

**DATA SET:** [meatspec.csv](#)

**HIPERPARAMÉTROS:**

ALPHA (LASSO – RIDGE)

NLAMBDA (ÉPOCAS – ITERACIONES PARA C-VALIDATION)

**LIBRERÍAS PARA APLICAR TÉCNICAS DE  
REGULARIZACIÓN**

scikit-learn



¿PREGUNTAS?

¡GRACIAS!