L2 REGULARIZED (RIDGE) REGRESSION

WHEN FITTING A MODEL,
WE CAN "SHRINK" COEFFICIENTS
TOWARDS ZERO

SHRINKING LESS IMPORTANT
OR REDUNDANT
FEATURE'S CONTRIBUTIONS

VIA

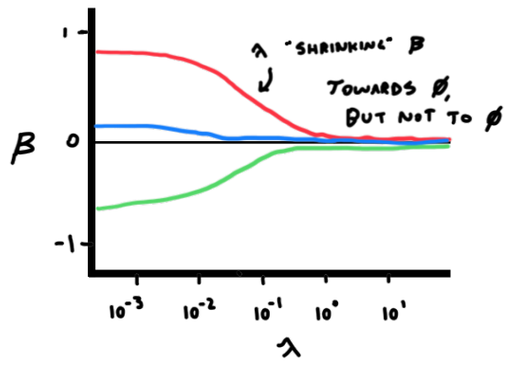
$$\lambda \cdot \sum \beta_j^2$$

PENALTY ADDED TO O.L.S. LOSS FUNCTION

LAMDA
MODULATES THE STRENGTH OF SHRINKAGE

SUM OF SQUARED COEFFICIENTS

eg. FOR 3 FEATURES

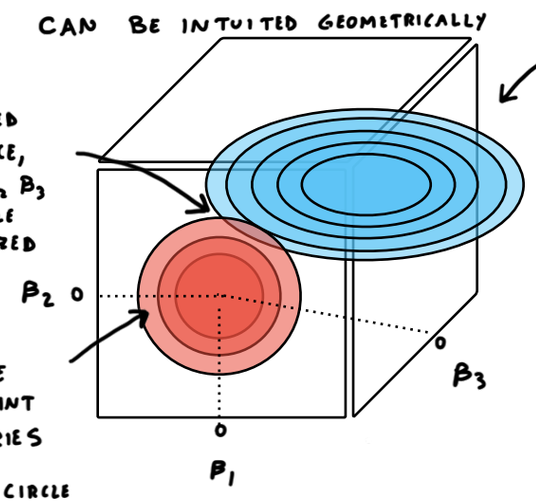


RIDGE REGULARIZATION
CAN BE INTUITED GEOMETRICALLY

RESIDUAL SUM OF SQUARES (RSS) CONTOURS

SMALLER CIRCLE = BETTER MODEL FIT

MODEL IS SELECTED AT THE INTERFACE, SUCH THAT β_1 β_2 β_3 ARE SHRUNK WHILE RSS IS MINIMIZED



RIDGE CONSTAINT BOUNDARIES

SMALLER CIRCLE = LARGER ↑

L1 REGULARIZED (LASSO) REGRESSION

WHEN FITTING A MODEL,
WE CAN "SHRINK" COEFFICIENTS
TO ZERO

"FEATURE SELECTION"

SHRINKING AND/OR ELIMINATING
LESS IMPORTANT
OR REDUNDANT
FEATURE'S CONTRIBUTIONS

VIA

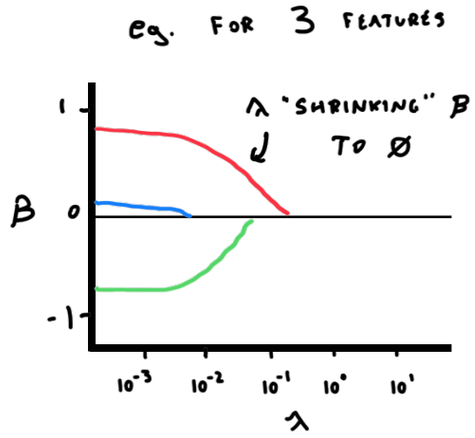
$$\lambda \cdot \sum |\beta_j|$$

PENALTY ADDED TO O.L.S. LOSS FUNCTION

LAMDA
MODULATES THE STRENGTH OF SHRINKAGE

SUM OF ABSOLUTE COEFFICIENTS

eg. FOR 3 FEATURES

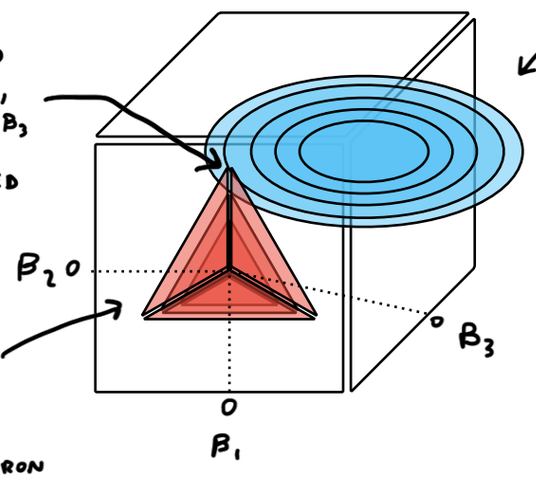


LASSO REGULARIZATION
CAN BE INTUITED GEOMETRICALLY

RESIDUAL SUM OF SQUARES (RSS) CONTOURS

SMALLER CIRCLE = BETTER MODEL FIT

MODEL IS SELECTED AT THE INTERFACE, SUCH THAT β_1 β_2 β_3 ARE SHRUNK WHILE RSS IS MINIMIZED



LASSO CONSTAINT BOUNDARIES

SMALLER POLYHEDRON = LARGER ↑

LASSO HAS A TENDENCY TO SELECT A "CORNER" OF THE POLYHEDRON, INDICATING THE SHRINKAGE OF SOME β TO Ø