NTHU Introduction to Machine learning (4602) Supplementary

Ridge Regression (Regression with L2 norm):

$L(W) = ||Y - XW||^2 + \lambda||W||^2$

Where:

L(W) is the objective function to be minimized.

Y is the vector of observed target values.

X is the design matrix of input features.

W is the vector of regression coefficients to be estimated.

λ (lambda) is the regularization parameter, which controls the strength of regularization.

The goal is to find the W values that minimize this objective function.

Calculate the derivative of L(W) with respect to W:

$\nabla L(W) = 2X^T(XW - Y) + 2\lambda W$

Set the derivative equal to zero and solve for W:

$2X^T(XW - Y) + 2\lambda W = 0$

$X^TXW - X^TY + \lambda W = 0$

$(X^TX + \lambda I)W = X^TY$

$W = (X^TX + \lambda I)^{-1}X^TY$

Where:

W is the vector of estimated regression coefficients.

X^T is the transpose of the design matrix.

I is the identity matrix.

Lasso Regression (Regression with L1 norm):

$L(W) = ||Y - XW||^2 + \lambda||W||_1$

Where:

L(W) is the objective function to be minimized.

Y is the vector of observed target values.

X is the design matrix of input features.

W is the vector of regression coefficients to be estimated.

λ (lambda) is the regularization parameter, which controls the strength of regularization.

To find the closed-form solution, we'll use subdifferential or subgradient.

Calculate the subdifferential of the L1 norm:

$d||W||_1 = \{ \text{sign}(W) \text{ for } W \neq 0, \text{any value in } [-1, 1] \text{ for } W = 0 \}$, where "sign(W)" returns the sign of each element in W.

Set up the subdifferential conditions for optimization:

$\nabla L(W) = 2X^T(XW - Y) + \lambda d||W||_1$

Find the subgradient of the L1 regularization term:

$\lambda d||W||_1 = \lambda * \{ \text{sign}(W) \text{ for } W \neq 0, \text{any value in } [-1, 1] \text{ for } W = 0 \}$

Set the subgradient conditions equal to zero:

$2X^T(XW - Y) + \lambda * \{ \text{sign}(W) \text{ for } W \neq 0, \text{any value in } [-1, 1] \text{ for } W = 0 \} = 0$

$2X^T(XW - Y) = -\lambda * \{ \text{sign}(W) \text{ for } W \neq 0, \text{any value in } [-1, 1] \text{ for } W = 0 \}$

$W = (X^TX)^{-1}(X^TY - \lambda * \{ \text{sign}(W) \text{ for } W \neq 0, \text{any value in } [-1, 1] \text{ for } W = 0 \})$