RIDGE REGRESSION

LET US RECALL THE LINEAR REGRESSION MODEL

$$y = x\beta + e$$

β IS ESTIMATED SO THAT $e^T e$ IS MINIMIZED.

IN THIS CASE:

$$\hat{\beta} = (x^T x)^{-1} x^T y \quad ; \quad \text{IF } (x^T x)^{-1} \text{ EXISTS}$$

$(x^T x)^{-1}$ DOES NOT EXIST, IF THE COLUMNS ARE LINEARLY DEPENDENT. THIS CAUSES $x^T x$ TO BE SINGULAR.

IF $(x^T x)^{-1}$ DOES NOT EXIST, THEN $(x^T x + \lambda I)^{-1}$ MIGHT EXIST, WHERE I = IDENTITY MATRIX, AND $\lambda > 0$ IS A USER DEFINED PARAMETER. IN THIS CASE:

$$\hat{\beta}_R = (x^T x + \lambda I)^{-1} (x^T y)$$

- IT CAN BE SHOWN THAT $\hat{\beta}_R$ IS THE SOLUTION ~~of~~:

$$\hat{\beta}_R = \underset{\beta}{\text{ARG MIN}} \left[\|y - x\beta\|^2 + \lambda \|\beta\|^2 \right], \quad \text{WHERE } \lambda > 0$$

AND $\|\cdot\|$ IS THE EUCLIDEAN NORM OF A VECTOR.

- ALTERNATELY, THE ABOVE $\hat{\beta}_R$ CAN BE OBTAINED BY SOLVING:

MINIMIZE $J(\beta) = \|y - x\beta\|^2$

SUBJECT TO: $\|\beta\|^2 < \rho$

THE USER SELECTED VALUE ρ LIMITS THE SPACE IN WHICH β EXISTS. THUS $\hat{\beta}_R$ IS THE REGULARIZED LEAST SQUARE SOLUTION FOR THE LINEAR REGRESSION TASK.

GIVEN $\hat{\beta}_R = \text{ARG} \min_{\beta} \left[\|Y - X\beta\|^2 + \lambda \|\beta\|^2 \right]$; $\lambda > 0$

PROVE: $\hat{\beta}_R = (X^T X + \lambda I)^{-1} (X^T Y)$

PROOF: MINIMIZ $(Y - X\beta)^T (Y - X\beta) + \lambda \beta^T \beta$

TAKE DERIVATIVE WITH RESPECT TO β AND EQUATE THE RESULT TO 0.

$L = (Y^T - \beta^T X^T)(Y - X\beta) + \beta^T \beta \lambda$

$= Y^T Y - Y^T X\beta - \beta^T X^T Y + (\beta^T X^T)(X\beta) + \lambda \beta^T \beta$

$Y^T X \beta$ IS SCALAR. THEREFORE $Y^T X\beta = \beta^T X^T Y$

$L = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta + \lambda \beta^T \beta$

$\frac{\partial L}{\partial \beta} = -2X^T Y + X^T X\beta + X^T X\beta + 2\lambda\beta = 0$

THAT IS: $(X^T X + \lambda I)\hat{\beta}_R = X^T Y$

$\hat{\beta}_R = (X^T X + \lambda I)^{-1} (X^T Y)$ $\square$