1 Gauss - Markov Therom Proof

$$Var(d'y) > Var(c'\hat{\beta})$$

$$Var(d'y) = Var(d'y - c'\hat{\beta} + c'\hat{\beta})$$
$$= Var(d'y - c'\hat{\beta}) + Var(c'\hat{\beta}) + 2Cov(d'y - c'\hat{\beta}, c'\hat{\beta})$$

$$Var(d'y - c'\hat{\beta}) = Var(d'y - l'y)$$
$$= Var((d' - l')y)$$
$$= Var((d - l)'y)$$
$$= (d-l)' Var(y)(d-l)$$
$$= (d-l)'(\sigma^2 I)(d-l)$$
$$= \sigma^2 (d-l)' I (d-l)$$
$$= \sigma^2 (d-l)'(d-l) > 0 \text{ by } (1)$$

$$Cov(d'y - c'\hat{\beta}, c'\hat{\beta}) = Cov(d'y - l_y, l'y)$$
$$= Cov((d-l)'y, l'y)$$
$$= (d-l)' Var(y) l$$
$$= \sigma^2 (d-l)' l$$
$$= \sigma^2 (d-l)' X [(X'X)^-]' c = 0 \text{ by } (2)$$

$$Var(d'y) = Var(d'y - c'\hat{\beta}) + Var(c'\hat{\beta})$$

$$Var(d'y) > Var(c'\hat{\beta}) \quad \text{- QED}$$

taken from:
public.iastate.edu

2a)

$$X = \begin{array}{cccc} & x_0 & x_1 & x_2 \\ & \begin{bmatrix} 1 & 1 & 5 \\ 1 & 3 & 1 \\ 1 & 3 & 5 \\ 1 & 5 & 1 \end{bmatrix} \end{array}$$

$$\hat{y}_{ij} = \frac{\langle z_i , z_j \rangle}{z_1 , z_1}$$

$$\hat{y}_{01} = 1 + 3 + 3 + 5 / 1 + 1 + 1 + 1 = 3$$

$$z_1 = \sum_{k=0} \hat{y}_{01} (z_0) = \begin{bmatrix} 1 \\ 3 \\ 3 \\ 5 \end{bmatrix} - 3 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 0 \\ 2 \end{bmatrix}$$

$$\hat{y}_{02} = 5 + 1 + 5 + 1 \;/\; 1 + 1 + 1 + 1 = 3$$

$$\hat{y}_{12} = -1$$

$$z_2 = \begin{bmatrix} 5 \\ 1 \\ 5 \\ 1 \end{bmatrix} - \left(3 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \left(-1 \begin{bmatrix} -2 \\ 0 \\ 0 \\ 2 \end{bmatrix} \right) \right) = \begin{bmatrix} 0 \\ -2 \\ 2 \\ 0 \end{bmatrix}$$

$$Z = \begin{bmatrix} 1 & -2 & 0 \\ 1 & 0 & -2 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{bmatrix}$$

b)

$$R = \begin{bmatrix} 1 & 3 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \qquad Z = \begin{bmatrix} 1 & -2 & 0 \\ 1 & 0 & -2 \\ 1 & 0 & 2 \\ 1 & -2 & 0 \end{bmatrix}$$

$X = Z\Gamma$

$$\begin{bmatrix} 1 & -2 & 0 \\ 1 & 0 & -2 \\ 1 & 0 & -22 \\ 1 & -2 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 3 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 5 \\ 1 & 3 & 1 \\ 1 & 3 & 5 \\ 1 & 5 & 1 \end{bmatrix}$$

4 $RSS(\lambda) = (Y - X\beta)^T(Y - X\beta) + \lambda\beta^T\beta$

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

$= Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta + \lambda\beta^T\beta$

$= \dfrac{\partial(Y^TY - 2(\beta^Tx^TY) + \beta^TX^TX\beta^T + \lambda\beta^T\beta)}{\partial\beta}$

$= -2(X^TY) + (2X^TX\beta) + 2\lambda\beta$

$2(X^TY) = (2X^TX\beta) + 2\lambda\beta$

$\dfrac{2(X^TY)}{2} = \dfrac{2\beta(X^TX + \lambda I)}{2}$

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

$\dfrac{X^TY}{(X^TX + \lambda I)} = \dfrac{\beta(X^TX + \lambda I)}{(X^TX + \lambda I)}$

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

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

5 $X\hat{\beta}^{ridge} = X(X^TX + \lambda I)^{-1}X^TY$

$= UD(D^2 + \lambda I)^{-1}DU^TY$

where U is a $N \times p$ orthogonal matrix with column U spanning the column space of X. D is a $p \times p$ diagonal matrix, with diagonal entries $d_1 \geq d_2 \geq \dots \geq d_p \geq 0$

$X\hat{\beta}^{ridge} = \sum_{j=1}^{p} u_j \dfrac{d_j^2}{d_j^2 + \lambda} u_j^T y$ ~~where the~~ where the u_j columns are the columns of U. Since $\lambda \geq 0$, $\dfrac{d_j^2}{(d_j^2 + \lambda)} \geq 1$. Ridge regression computes to coordinates of y with respect to the orthonormal basis U. The coordinates shrink by the factors $\dfrac{d_j^2}{(d_j^2 + \lambda)}$. A greater shrinkage are applied to the smaller d_j^2