DS 3001: Linear Models

$\hat{y}_i = b_0 + b_1 z_{i1} + b_2 z_{i2}$

$\frac{1}{N}\sum_{i=1}^{N} z_{ij} = 0$

1.1)

$SSE = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$

$= \sum_{i=1}^{N} (y_i - (b_0 + b_1 z_{i1} + b_2 z_{i2}))^2$

$\boxed{SSE = \sum_{i=1}^{N} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2})^2}$

1.2)

$\frac{dSSE}{db_0} = \sum_{i=1}^{N} \frac{d}{db_0} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2})^2$

$= -2 \sum_{i=1}^{N} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2})$

$= -2 \sum_{i=1}^{N} e_i$

$\frac{dSSE}{db_1} = \sum_{i=1}^{N} \frac{d}{db_1} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2})^2$

$= -2 \sum_{i=1}^{N} z_{i1} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2})$

$= -2 \sum_{i=1}^{N} z_{i1} e_i$

$\frac{dSSE}{db_2} = \sum_{i=1}^{N} \frac{d}{db_2} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2})^2$

$= -2 \sum_{i=1}^{N} z_{i2} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2})^2$

$= -2 \sum_{i=1}^{N} z_{i2} e_i$

1.3)

normal equations: $\sum_{i=1}^{N} e_i = 0$, $\sum_{i=1}^{N} e_i z_{i1} = 0$, $\sum_{i=1}^{N} e_i z_{i2} = 0$

dividing by N, $\frac{1}{N}\sum_{i=1}^{N} e_i = 0$ average residual is zero.

other two equations:

$\sum_{i=1}^{N} e_i z_{i1} = 0 \rightarrow e \cdot z_1 = 0$, $\sum_{i=1}^{N} e_i z_{i2} = 0 \rightarrow e \cdot z_2 = 0$

residual vector is orthogonal, just like in single linear regression

1.4) $b_0^* = \bar{y}$ ← optimal intercept

$\sum_{i=1}^{N} e_i = 0$

$\sum_{i=1}^{N} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2}) = 0$

$\sum_{i=1}^{N} y_i - N b_0 - b_1 \sum_{i=1}^{N} z_{i1} - b_2 \sum_{i=1}^{N} z_{i2} = 0$

$\sum_{i=1}^{N} z_{i1} = 0, \sum_{i=0}^{N} z_{i2} = 0$ ← because predictors are centered

$\sum_{i=1}^{N} y_i - N b_0 = 0 \rightarrow \sum_{i=1}^{N} y_i = N b_0$

$b_0^* = \frac{1}{N} \sum_{i=1}^{N} y_i = \bar{y}$ ← optimal intercept is sample mean of y

1.5)

$A_b = C$

$$\begin{bmatrix} \sum_{i=1}^{N} z_{i1}^2 & \sum_{i=1}^{N} z_{i1} z_{i2} \\ \sum_{i=1}^{N} z_{i1} z_{i2} & \sum_{i=1}^{N} z_{i1}^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} z_{i1}(y_i - \bar{y}) \\ \sum_{i=1}^{N} z_{i2}(y_i - \bar{y}) \end{bmatrix}$$

1.6) $z_{ij} = x_{ij} - m_j$

$$\frac{1}{N}\begin{bmatrix} \sum_{i=1}^{N} z_{i1}^2 & \sum_{i=1}^{N} z_{i1} z_{i2} \\ \sum_{i=1}^{N} z_{i1} z_{i2} & \sum_{i=1}^{N} z_{i1}^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \frac{1}{N}\begin{bmatrix} \sum_{i=1}^{N} z_{i1}(y_i - \bar{y}) \\ \sum_{i=1}^{N} z_{i2}(y_i - \bar{y}) \end{bmatrix}$$

following equation at top of question, we substitute and each entry of A/N is the covariance matrix of the original predictors

For c,

$\frac{1}{N} \sum_{i=1}^{N} z_{i1}(y_i - \bar{y}) = \mathrm{cov}(z_1, y) = \mathrm{cov}(x_1, y)$

c is the vector of covariances between each predictor and response.

This tells us that intuitively, we're aligning the joint variability of the predictors with their joint variability with the outcome.