Homework 2: Multiple Linear Regression

Introduction to Machine Learning
Homework 2: Multiple Linear Regression

Shangwen Yan | N17091204 | sy2160

1.

(a) past sales

(b) extract numeric score as vector x_1 , frequency of occurence of words as vector x_2 (positive when good, negative when bad, and assign weight towards different words), the model is like:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

(c) $\hat{y} = \beta_0 + \beta_1 (x_{11} + \frac{1}{2} x_{12}) + \beta_2 x_2$, numeric score from 1 to 5 as x_{11} , numeric score from 1 to 10 as x_{12}

(d) score from 1 to 5 as x_{11} ; rating as x_{12} , $x_{12} = 5$ when good, $x_{12} = 1$ when bad; no numeric rating as $x_{13} = 2.5$; so, $x_1 = x_{11} + x12 + x13$ and the model is still: $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

(e) I would suggest fraction of reviews with the word "good", but the total number of reviews also have significance(more popular, more reviews), so I would also suggest to add another predictor of the total number of reviews(both good and bad)

2.

(a) assume the model is:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

(b)

$$\hat{y} = 0.75 + 2.5 x_1 + 3.5 x_2$$

3.

(a) M+N+1

(b)

$$\mathbf{y} = \mathbf{A}\beta + \varepsilon$$

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{T-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & x_0 & 0 & 0 & \cdots \\ y_0 & 0 & \cdots & x_1 & x_0 & 0 & \cdots \\ y_1 & y_0 & \cdots & x_2 & x_1 & x_0 & \cdots \\ y_2 & y_1 & \cdots & x_3 & x_2 & x_1 & \cdots \\ \vdots & \vdots & \ddots & \vdots & & & \\ y_{T-2} & \cdots & y_{T-1-M} & x_{T-1} & x_{T-2} & \cdots & x_{T-1-N} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_M \\ b_0 \\ \vdots \\ b_N \end{bmatrix} + \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_{T-1} \end{bmatrix}$$

(c)

$$\frac{1}{T}(A^Ty)_{i,j} = \frac{1}{T}\begin{bmatrix} \sum_{k=0}^{T-1-1} y_k y_{k+1} \\ \vdots \\ \sum_{k=0}^{T-1-l} y_k y_{k+l} \\ \vdots \\ \sum_{k=0}^{T-1-M} y_k y_{k+M} \\ \sum_{k=0}^{T-1-0} x_k y_{k+0} \\ \vdots \\ \sum_{k=0}^{T-1-l} x_k y_{k+l} \\ \vdots \\ \sum_{k=0}^{T-1-N} x_k y_{k+N} \end{bmatrix} \overset{M\ll,N\ll T}{\approx} \frac{1}{T}\begin{bmatrix} \sum_{k=0}^{T-1} y_k y_{k+1} \\ \vdots \\ \sum_{k=0}^{T-1} y_k y_{k+l} \\ \vdots \\ \sum_{k=0}^{T-1} y_k y_{k+M} \\ \sum_{k=0}^{T-1} x_k y_{k+0} \\ \vdots \\ \sum_{k=0}^{T-1} x_k y_{k+l} \\ \vdots \\ \sum_{k=0}^{T-1} x_k y_{k+N} \end{bmatrix} = \frac{1}{T}\begin{bmatrix} R_{yy}(1) \\ R_{yy}(2) \\ \vdots \\ R_{yy}(M) \\ R_{xy}(0) \\ R_{xy}(1) \\ \vdots \\ R_{xy}(N) \end{bmatrix} = \begin{cases} \frac{1}{T}R_{yy}(i-j+1), i \le M \\ \frac{1}{T}R_{xy}(i-j-M)\ , M \le i \le M+N \end{cases}$$

with same derivative process, we can get:

$$\frac{1}{T}(A^TA)_{i,j} = \frac{1}{T}\begin{bmatrix} R_{yy}(0) & R_{yy}(1) & \cdots & R_{yy}(M-1) & R_{xy}(-1) & \cdots & R_{xy}(N-1) \\ R_{yy}(1) & R_{yy}(0) & & & & & \\ \vdots & \ddots & & & & & \\ R_{yy}(M-1) & & & & & & \\ R_{xy}(-1) & & & & & & \\ R_{xy}(0) & & & & & & \\ \vdots & & & & & & \\ R_{xy}(N-1) & R_{xy}(N-2) & \cdots & & & & \end{bmatrix}$$

$$= \frac{1}{T}\begin{bmatrix} R_{yy}(i-j) & R_{xy}(i-j-M-1) \\ R_{xy}(i-j-M-1) & R_{xx}(i-j) \end{bmatrix}$$

4.

(a)

$$\mathbf{x} = \mathbf{A}\beta$$

$$\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_k \\ \vdots \\ x_{N-1} \end{bmatrix} = \begin{bmatrix} \cos(\Omega_1 \times 0) & \cdots & \cos(\Omega_l \times 0) & \sin(\Omega_1 \times 0) & \cdots & \sin(\Omega_l \times 0) \\ \cos(\Omega_1 \times 1) & \cdots & \cos(\Omega_l \times 1) & \sin(\Omega_1 \times 1) & \cdots & \sin(\Omega_l \times 1) \\ \vdots & \vdots & \vdots & \vdots & & \\ \cos(\Omega_1 \times (N-1)) & \cdots & \cos(\Omega_l \times (N-1)) & \sin(\Omega_1 \times (N-1)) & \cdots & \sin(\Omega_l \times (N-1)) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_l \\ b_1 \\ \vdots \\ b_l \end{bmatrix}$$

we can compute:

$$R_{AA} = \frac{1}{N-1}A^TA$$

$$R_{Ax} = \frac{1}{N-1} A^T x$$

$$\beta = R_{AA}^{-1} R_{Ax} = (A^T A)^{-1} A^T x$$

(b) No, because we don't have predictors