Solution to Homework 2

Name: Chen Shen

NetID: cs5236

1.

(a) For example, the sales of a product.

(b) Let x_i be the frequency of occurrence of a certain word and y be the sales of the product. Then we have the following linear model

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

(c) Normalized the scores so that they have a common limit.

(d) We can use an array of one-hot coding, such as [score, good, bad, no rating].

(e) Of course the fraction of reviews with the word "good". We need to consider both the reviews with the word "good" and the total number of reviews.

2.

(a)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

(b) First, we can have a formula in matrice form

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

where

$$\mathbf{y} = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 7 \end{bmatrix}, \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

So we have

$$\beta = \begin{bmatrix} 0.75 \\ 2.5 \\ 3.5 \end{bmatrix},$$

i.e. $\beta_0 = 0.75$, $\beta_1 = 2.5$, $\beta_2 = 3.5$.

3.

(a) The vactor β could be as follows. There are $M + N + 1$ unknown parameters.

$$\beta = [a_1, a_2, \cdots, a_M, b_0, b_1, \cdots, b_N]^{\mathrm{T}}$$

(b) The matrix $\mathbf{A}$ and $\mathbf{y}$ could be as follows.

$$\mathbf{A} = \begin{bmatrix} y_{M-1} & \cdots & y_0 & x_M & \cdots & x_{M-N} \\ y_M & \cdots & y_1 & x_{M+1} & \cdots & x_{M-N+1} \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ y_{T-2} & \cdots & y_0 & x_M & \cdots & x_{T-N-1} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_M \\ \vdots \\ y_{T-1} \end{bmatrix}.$$

(c) First, we can divide the matrix $\mathbf{A}$ into two parts.

$$\mathbf{A} = \begin{bmatrix} \mathbf{A_y} & \mathbf{A_x} \end{bmatrix},$$

where

$$\mathbf{A_y} = \begin{bmatrix} y_{M-1} & \cdots & y_0 \\ y_M & \cdots & y_1 \\ \vdots & \cdots & \vdots \\ y_{T-2} & \cdots & y_0 \end{bmatrix}, \ \mathbf{A_x} = \begin{bmatrix} x_M & \cdots & x_{M-N} \\ x_{M+1} & \cdots & x_{M-N+1} \\ \vdots & \cdots & \vdots \\ x_M & \cdots & x_{T-N-1} \end{bmatrix}.$$

So we have

$$\begin{aligned} \frac{1}{T}\mathbf{A}^{\mathrm{T}}\mathbf{A} &= \frac{1}{T}\begin{bmatrix} \mathbf{A_y^{\mathrm{T}}} \\ \mathbf{A_x^{\mathrm{T}}} \end{bmatrix} \begin{bmatrix} \mathbf{A_y} & \mathbf{A_x} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{T}\mathbf{A_y^{\mathrm{T}}}\mathbf{A_y} & \frac{1}{T}\mathbf{A_y^{\mathrm{T}}}\mathbf{A_x} \\ \frac{1}{T}\mathbf{A_x^{\mathrm{T}}}\mathbf{A_y} & \frac{1}{T}\mathbf{A_x^{\mathrm{T}}}\mathbf{A_x} \end{bmatrix} \end{aligned}$$

Now focus on a certain element.

$$\begin{aligned} \frac{1}{T}(\mathbf{A_y^{\mathrm{T}}}\mathbf{A_y})_{i,j} &= \sum_{k=1}^{T-M} (\mathbf{A_y^{\mathrm{T}}})_{i,k}(\mathbf{A_y})_{k,j} \\ &= \sum_{k=1}^{T-M} (\mathbf{A_y})_{k,i}(\mathbf{A_y})_{k,j} \\ &= \sum_{k=1}^{T-M} y_{M+k-i-1}y_{M+k-j-1} \\ &= \sum_{k=M-i}^{T-i-1} y_k y_{k+(i-j)} \end{aligned}$$

Since $T \gg N$ and $T \gg M$, it goes like

$$\frac{1}{T}(\mathbf{A_y^{\mathrm{T}}}\mathbf{A_y})_{i,j} \approx \sum_{k=0}^{T-1} y_k y_{k+(i-j)} = R_{yy}(i-j).$$

Similarly, we have

$$\begin{aligned} \frac{1}{T}(\mathbf{A_x^{\mathrm{T}}}\mathbf{A_x})_{i,j} &\approx \sum_{k=0}^{T-1} x_k x_{k+(i-j)} = R_{xx}(i-j), \\ \frac{1}{T}(\mathbf{A_y^{\mathrm{T}}}\mathbf{A_x})_{i,j} = \frac{1}{T}(\mathbf{A_x^{\mathrm{T}}}\mathbf{A_y})_{i,j} &\approx \sum_{k=0}^{T-1} x_k y_{k+(i-j)} = R_{xy}(i-j). \end{aligned}$$

To $\frac{1}{T}\mathbf{A}^{\mathrm{T}}\mathbf{y}$, we can take the same steps. First,

$$\begin{aligned} \frac{1}{T}\mathbf{A}^{\mathrm{T}}\mathbf{y} &= \frac{1}{T}\begin{bmatrix} \mathbf{A_y^{\mathrm{T}}} \\ \mathbf{A_x^{\mathrm{T}}} \end{bmatrix}\mathbf{y} \\ &= \begin{bmatrix} \frac{1}{T}\mathbf{A_y^{\mathrm{T}}}\mathbf{y} \\ \frac{1}{T}\mathbf{A_x^{\mathrm{T}}}\mathbf{y} \end{bmatrix}. \end{aligned}$$

Then concentrate on one element.

$$\begin{aligned}
\frac{1}{T}(\mathbf{A}_{\mathbf{y}}^{\mathrm{T}}\mathbf{y})_{i,1} &= \sum_{k=1}^{T-M} (\mathbf{A}_{\mathbf{y}}^{\mathrm{T}})_{i,k}(\mathbf{y})_{k,1} \\
&= \sum_{k=1}^{T-M} (\mathbf{A}_{\mathbf{y}})_{k,i}(\mathbf{y})_{k,1} \\
&= \sum_{k=1}^{T-M} y_{M+k-i-1}y_{M+k-1} \\
&= \sum_{k=M-i}^{T-i-1} y_k y_{k+i} \\
&\approx \sum_{k=0}^{T-1} y_k y_{k+i} \\
&= R_{yy}(i)
\end{aligned}$$

And also we can get

$$\begin{aligned}
\frac{1}{T}(\mathbf{A}_{\mathbf{x}}^{\mathrm{T}}\mathbf{y})_{i,1} &\approx \sum_{k=0}^{T-1} x_k y_{k+i} \\
&= R_{xy}(i)
\end{aligned}$$

In conclusion, $\frac{1}{T}\mathbf{A}^{\mathrm{T}}\mathbf{A}$ and $\frac{1}{T}\mathbf{A}^{\mathrm{T}}\mathbf{y}$ can be approximately computed from the autocorrelation functions.

4.

(a) We can define

$$\mathbf{A} = \begin{bmatrix} cos(\Omega_1(0)) & \cdots & cos(\Omega_L(0)) & sin(\Omega_1(0)) & \cdots & sin(\Omega_L(0)) \\ cos(\Omega_1(1)) & \cdots & cos(\Omega_L(1)) & sin(\Omega_1(1)) & \cdots & sin(\Omega_L(1)) \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ cos(\Omega_1(T-1)) & \cdots & cos(\Omega_L(T-1)) & sin(\Omega_1(T-1)) & \cdots & sin(\Omega_L(T-1)) \end{bmatrix}$$

and

$$\mathbf{x} = \begin{bmatrix} x_0 \\ \vdots \\ x_{T-1} \end{bmatrix}, \beta = \begin{bmatrix} a_1 \\ \vdots \\ a_L \\ b_1 \\ \vdots \\ b_L \end{bmatrix}.$$

Then we have $\mathbf{x} \approx \mathbf{A}\beta$.

(b) No, the model is nonlinear.