1.

a) The target value should be the past sales.

b) The more frequent the positive word occur, the more likely it could be sold. Therefore, it is possible that frequency of positive word can linearly be relative to sales.

c) The model will be inaccurate. For example, a score 1 with range from 1 to 5 should have the same effect as a score 2 with range from 1 to 10. However, the model would consider the second score being greater than the first one.

d) I would normalize them by following steps: For (a), keep it as it is. For (b), score good rating 5, and bad rating 1. For (c), give score 2.5 for each.

e) I would choose (b) since the total number of reviews vary, which would make result using (a) inaccurate.

2.

a) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

b) $y = 0.75 + 2.5x_1 + 3.5x_2$

3.

a) $\boldsymbol{\beta} = [a_1, a_2],\ \boldsymbol{x} = [x_1 e^{-x_1-x_2}, x_2 e^{-x_1-x_2}]$

$\boldsymbol{\beta}^T \phi(\boldsymbol{x}) = a_1 x_1 e^{-x_1-x_2} + a_2 x_2 e^{-x_1-x_2} = (a_1 x_1 + a_2 x_2) e^{-x_1-x_2} = \hat{y}$

b) $\boldsymbol{\beta} = \begin{cases} [a_1, a_2] \text{ if } x < 1 \\ [a_3, a_4] \text{ if } x \geq 1 \end{cases},\ \boldsymbol{x} = [1 + x]$

$\boldsymbol{\beta}^T \phi(\boldsymbol{x}) = \begin{cases} a_1 + a_2 x \text{ if } x < 1 \\ a_3 + a_4 x \text{ if } x \geq 1 \end{cases} = \hat{y}$

c) $\boldsymbol{\beta} = [e^{a_2}, a_1 e^{a_2}],\ \boldsymbol{x} = [e^{-x_2}, x_1 e^{-x_2}]$

$\boldsymbol{\beta}^T \phi(\boldsymbol{x}) = e^{a_2} e^{-x_2} + a_1 e^{a_2} x_1 e^{-x_2} = e^{-x_2+a_2} + a_1 x_1 e^{-x_2+a_2}$

$= (1 + a_1 x_1) e^{-x_2+a_2} = \hat{y}$

4.

a) $\boldsymbol{\beta} = \begin{bmatrix} a_1 \\ \vdots \\ a_M \\ b_0 \\ \vdots \\ b_N \end{bmatrix}$

There are $M + N + 1$ unknown parameters.

b) $\boldsymbol{A} = \begin{bmatrix} y_{M-1} & y_{M-2} & \cdots & y_0 & x_{M-1} & x_{M-2} & \cdots & x_0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{k-1} & y_{k-2} & \cdots & y_{k-M} & x_k & x_{k-1} & \cdots & x_{k-N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{T-2} & y_{T-3} & \cdots & y_{T-1-M} & x_{T-1} & x_{T-2} & \cdots & x_{T-1-N} \end{bmatrix}$ assume $M > N + 1$.

c) Let $A_y = \begin{bmatrix} y_{M-1} & \cdots & y_0 \\ \vdots & \vdots & \vdots \\ y_{T-2} & \cdots & y_{T-1-M} \end{bmatrix}$, $A_x = \begin{bmatrix} x_{M-1} & \cdots & x_0 \\ \vdots & \vdots & \vdots \\ x_{T-1} & \cdots & x_{T-1-N} \end{bmatrix}$.

Then $\boldsymbol{A}$ can be represented as $\boldsymbol{A} = [A_y \quad A_x]$.

$$A_x^T A_y = \begin{bmatrix} x_0 y_{M-1} + \cdots + x_{T-1-N} y_{T-2} & \cdots & x_0 y_0 + \cdots + x_{T-1-N} y_{T-1-M} \\ \vdots & \vdots & \vdots \\ x_{M-1} y_{M-1} + \cdots + x_{T-1} y_{T-2} & \cdots & x_{M-1} y_0 + \cdots + x_{T-1} y_{T-1-M} \end{bmatrix}$$
$$\approx T \begin{bmatrix} R_{xy}(M-1) & \cdots & R_{xy}(0) \\ \vdots & \ddots & \vdots \\ R_{xy}(0) & \cdots & R_{xy}(1-M) \end{bmatrix}$$

$$A_x^T A_x = \begin{bmatrix} x_0 x_{M-1} + \cdots + x_{T-1-N} x_{T-2} & \cdots & x_0 x_0 + \cdots + x_{T-1-N} y_{T-1-N} \\ \vdots & \vdots & \vdots \\ x_{M-1} x_{M-1} + \cdots + x_{T-2} x_{T-2} & \cdots & x_{M-1} x_0 + \cdots + x_{T-1} x_{T-1-N} \end{bmatrix}$$
$$\approx T \begin{bmatrix} R_{xx}(M-1) & \cdots & R_{xx}(0) \\ \vdots & \ddots & \vdots \\ R_{xx}(0) & \cdots & R_{xx}(1-M) \end{bmatrix}$$

$$A_y^T A_y = \begin{bmatrix} y_0 y_{M-1} + \cdots + y_{T-1-M} y_{T-2} & \cdots & y_0 y_0 + \cdots + y_{T-1-M} y_{T-1-M} \\ \vdots & \vdots & \vdots \\ y_{M-1} y_{M-1} + \cdots + y_{T-2} y_{T-2} & \cdots & y_{M-1} y_0 + \cdots + y_{T-2} y_{T-1-M} \end{bmatrix}$$
$$\approx T \begin{bmatrix} R_{yy}(M-1) & \cdots & R_{yy}(0) \\ \vdots & \ddots & \vdots \\ R_{yy}(0) & \cdots & R_{yy}(1-M) \end{bmatrix}$$

$$A_y^T A_x = \begin{bmatrix} y_0 x_{M-1} + \cdots + y_{T-1-M} x_{T-2} & \cdots & y_0 x_0 + \cdots + y_{T-1-M} x_{T-1-N} \\ \vdots & \vdots & \vdots \\ y_{M-1} x_{M-1} + \cdots + y_{T-2} x_{T-2} & \cdots & y_{M-1} x_0 + \cdots + y_{T-2} x_{T-1-N} \end{bmatrix}$$
$$\approx T \begin{bmatrix} R_{xy}(1-M) & \cdots & R_{xy}(0) \\ \vdots & \ddots & \vdots \\ R_{xy}(0) & \cdots & R_{xy}(M-1) \end{bmatrix}$$

$$\frac{1}{T}A^TA = \frac{1}{T}\begin{bmatrix} A_x^TA_y & A_x^TA_x \\ A_y^TA_y & A_y^TA_x \end{bmatrix}$$

$$\approx \begin{bmatrix} R_{xy}(M-1) & \cdots & R_{xy}(0) & R_{xx}(M-1) & \cdots & R_{xx}(0) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ R_{xy}(0) & \cdots & R_{xy}(1-M) & R_{xx}(0) & \cdots & R_{xx}(1-M) \\ R_{yy}(M-1) & \cdots & R_{yy}(0) & R_{xy}(1-M) & \cdots & R_{xy}(0) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ R_{yy}(0) & \cdots & R_{yy}(1-M) & R_{xy}(0) & \cdots & R_{xy}(M-1) \end{bmatrix}$$

$$\frac{1}{T}A^Ty = \frac{1}{T}\begin{bmatrix} x_0y_0 + \cdots x_{T-1-N}y_{T-1-N} \\ \vdots \\ x_{M-1}y_0 + \cdots + x_{T-1}y_{T-M-2} \\ y_0y_0 + \cdots + y_{T-1-M}y_{T-1-M} \\ \vdots \\ y_{M-1}y_0 + \cdots + y_{T-2}y_{T-M-3} \end{bmatrix} \approx \begin{bmatrix} R_{xy}(0) \\ \vdots \\ R_{xy}(M-1) \\ R_{yy}(0) \\ \vdots \\ R_{yy}(M-1) \end{bmatrix}$$

5.

a) $\boldsymbol{\beta} = \begin{bmatrix} a_1 \\ \vdots \\ a_L \\ b_1 \\ \vdots \\ b_L \end{bmatrix}$

$$\boldsymbol{A} = \begin{bmatrix} \cos\Omega_1 L & \cdots & \cos\Omega_L L & \sin\Omega_1 L & \cdots & \sin\Omega_L L \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cos\Omega_1 k & \cdots & \cos\Omega_L k & \sin\Omega_1 k & \cdots & \sin\Omega_L k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cos\Omega_1(N-1) & \cdots & \cos\Omega_L(N-1) & \sin\Omega_1(N-1) & \cdots & \sin\Omega_L(N-1) \end{bmatrix}$$

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

$$= \begin{bmatrix} a_1\cos\Omega_1 L + \cdots + a_L\cos\Omega_L L + b_1\sin\Omega_1 L + \cdots + b_L\sin\Omega_L L \\ \vdots \\ a_1\cos\Omega_1(N-1) + \cdots + a_L\cos\Omega_L(N-1) + b_1\sin\Omega_1(N-1) + \cdots + b_L\sin\Omega_L(N-1) \end{bmatrix}$$

$$\boldsymbol{x} = \begin{bmatrix} x_L \\ \vdots \\ x_{N-1} \end{bmatrix}$$

To solve a_l and b_l, we have $N-1-L$ equations. There are $2L$ a_l and b_l, thus we can find them by solving equations in $\boldsymbol{x}$ if $N \geq 3L+1$.

b) No. if Ω_l becomes unknown, x will be a function of $a_l, b_l, \cos\Omega_l$ and $\sin\Omega_l$, which is not linear.

6.

a)

```
x[:, 2] = x[:, 1] * x[:, 2]
yhat = x.dot(beta)
```

b)

```
x = np.reshape(x, (1, -1))
alpha = np.reshape(alpha, (1, -1))
beta = np.reshape(beta, (-1, 1))
m = np.dot(-beta, x)
yhat = alpha.dot(np.exp(m))
```

c)

```
x = np.expand_dims(x, 1).repeat(m, axis=1)
y = np.expand_dims(y, 0).repeat(n, axis=0)
dist = x - y
dist = dist ** 2
dist = dist.sum(axis=-1)
print(dist)
```