1.

(a)I think the target variable is sales.

(b)$y = \beta_0 + \beta_1 x_0 + \beta_1 x_1 + \beta_2 x_2$

In this formula, x_0 represents the feature of numeric score and x_1 represents the feature of the frequency of occurrence of positive words like "good", and x_2 represents the frequency of occurrence of negative words like"bad", "doesn't work".

(c)Use normalization $z = \frac{x-\mu}{\sigma}$, put every numeric score into a number from zero to one.

(d) First of all, I ignore the reviews which don't have numeric rating, and then I define the score from 0 to2 is belong to "bad" products, and score from 3 to5 is belong to "good" products.Finally, I can deal with (a) and (b) reviews at the same time.

(e)I will use fraction of reviews with the word "good"

3.

(a)$\beta = (a_1, a_2)^T$

$if\ \beta = (\beta_1, \beta_2)^T$,so $a_1 = \beta_1, a_2 = \beta_2$

$\phi((x_1, x_2)) = \left(x_1 e^{-x_1 - x_2}, x_2 e^{-x_1 - x_2}\right)^T$

(b) $\beta = (a_1\theta(t-1) + a_3\theta(-t+1), a_2\theta(t-1) + a_3\theta(-t+1))$

$$\theta(x) = \begin{Bmatrix} 0 & x \geq 0 \\ 1 & x < 0 \end{Bmatrix}$$

$$if\ \beta = \begin{Bmatrix} (\beta_1, \beta_2)^T\ x < 1 \\ (\beta_3, \beta_4)^T\ x \geq 1 \end{Bmatrix}$$ so $a_1 = \beta_1, a_2 = \beta_2, a_3 = \beta_3, a_4 = \beta_4$

$\emptyset(x) = (1, x)^T$

(c)$\beta = \left(e^{a_2}, a_1 e^{a_2}\right)^T$

if $\beta = (\beta_1, \beta_2)^T, so\ a_1 = \beta_2/\beta_1, a_2 = ln\beta_1$

$\phi\big((x_1, x_2)\big) = (e^{-x_2}, x_1 e^{-x_2})$

4.

(1)$\beta = (a_1, a_2, \cdots, a_M, b_0, b_1, b_2, \cdots, b_N)^T$

There are M+N+1 unknown parameters.

(2) $$A=\begin{bmatrix} 0 & 0 & \cdots & 0 & x_0 & 0 & \cdots & 0 \\ y_0 & 0 & \cdots & 0 & x_1 & x_0 & \cdots & 0 \\ \vdots & & & & \vdots & & & \vdots \\ y_{M-1} & & \cdots & y_0 & x_M & & \cdots & x_{M-N} \\ y_M & & \cdots & y_1 & x_{M+1} & & \cdots & x_{M-N+1} \\ \vdots & & & & & & & \\ y_{T-2} & & \cdots & y_{T-M-1} & x_{T-1} & & \cdots & x_{T-N-1} \end{bmatrix}$$

(3) $$\frac{1}{T}(A^TA)=\begin{bmatrix} R_{yy}(0) & R_{yy}(1) & \cdots & R_{yy}(M-1) & R_{xy}(-1) & R_{xy}(0) & \cdots & R_{xy}(N-1) \\ R_{yy}(-1) & R_{yy}(0) & \cdots & R_{yy}(M-2) & R_{xy}(-2) & & \cdots & R_{xy}(N-2) \\ \vdots & & & & & & & \\ R_{yy}(1-M) & \cdots & & R_{yy}(0) & R_{xy}(-M) & & \cdots & R_{xy}(N-M) \\ R_{xy}(1) & R_{xy}(2) & \cdots & R_{xy}(M) & R_{xx}(0) & & \cdots & R_{xx}(N) \\ R_{xy}(-1) & R_{xy}(0) & \cdots & R_{xy}(M-1) & R_{xx}(-1) & R_{xx}(0) & \cdots & R_{xx}(N-1) \\ \vdots & & & & & & & \\ R_{xx}(1-N) & \cdots & & R_{xy}(M-N) & R_{xx}(-N) & & \cdots & R_{xx}(0) \end{bmatrix}$$

$$\frac{1}{T}(A^Ty)=[R_{yy}(1)\ R_{yy}(2)\ \cdots\ R_{yy}(M)\ R_{xy}(0)\ \cdots\ R_{xy}(N)]^T.$$

6.

(a)

```
yhat=np.dot(X[:,:2],beta[:2]) + X[:,1]*X[:,2]*beta[2]
```

(b)

```
n=len(x)
m=len(alpha)
yhat=np.sum(np.exp(x * beta.T[:m]) * alpha.T, axis=1)
```

(c)

```
DXY=x[:,None,:]-y[None,:]
```

```
dist=np.sum(x**2, axis=1)*np.sum((y.T)**2,axis=0)-2*n.dot(y.T);
```