1.

(a)I think the target variable is sales.

(b)

In this formula, represents the feature of numeric score and represents the feature of the frequency of occurrence of positive words like “good”, and represents the frequency of occurrence of negative words like“bad”, “doesn’t work”.

(c)Use normalization , 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)

so

(b)

so

(c)

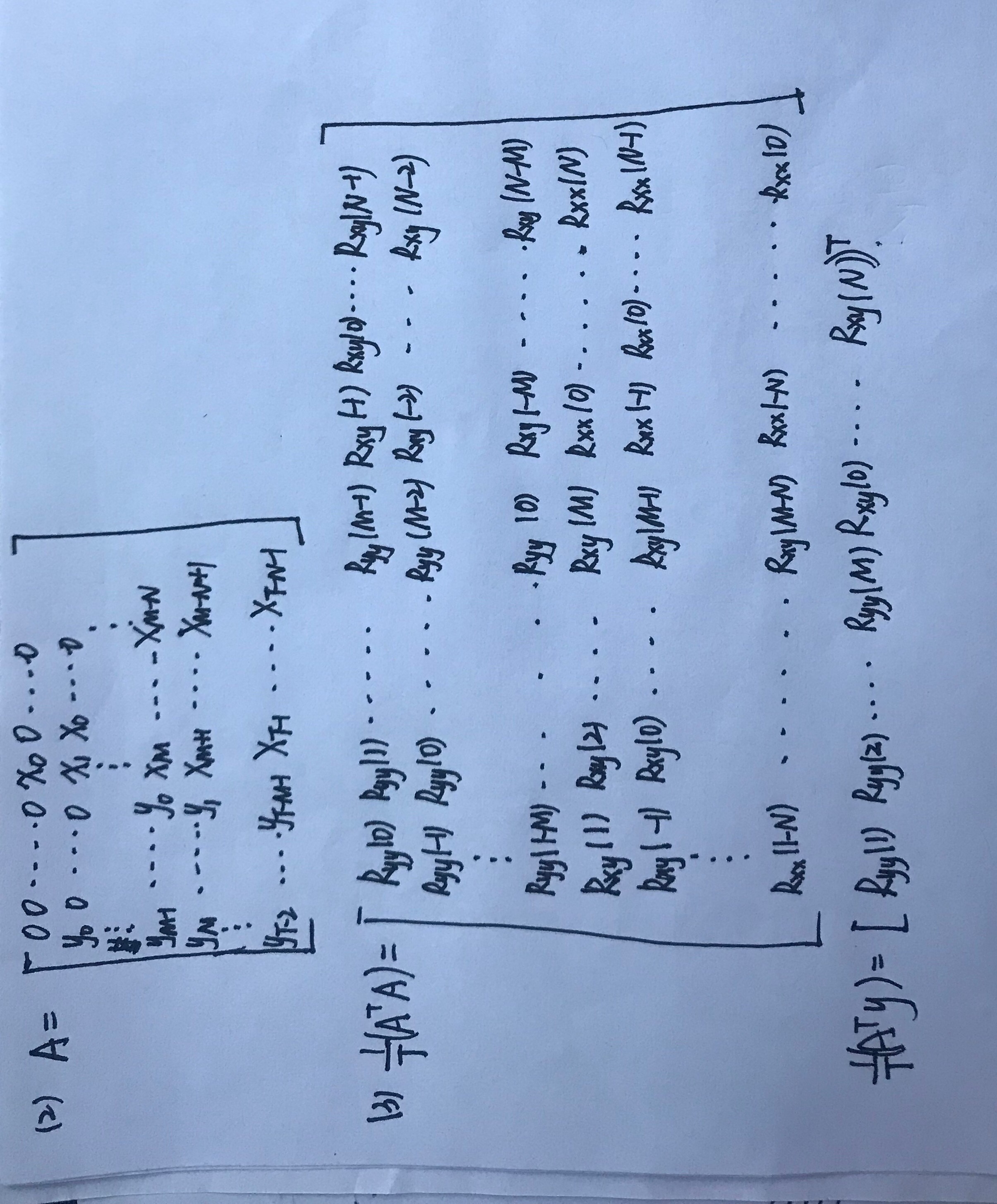
if

)

4.

(1)

There are M+N+1 unknown parameters.



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);