Homework 2

Name: Jingyi Lou N16071872 jl8166

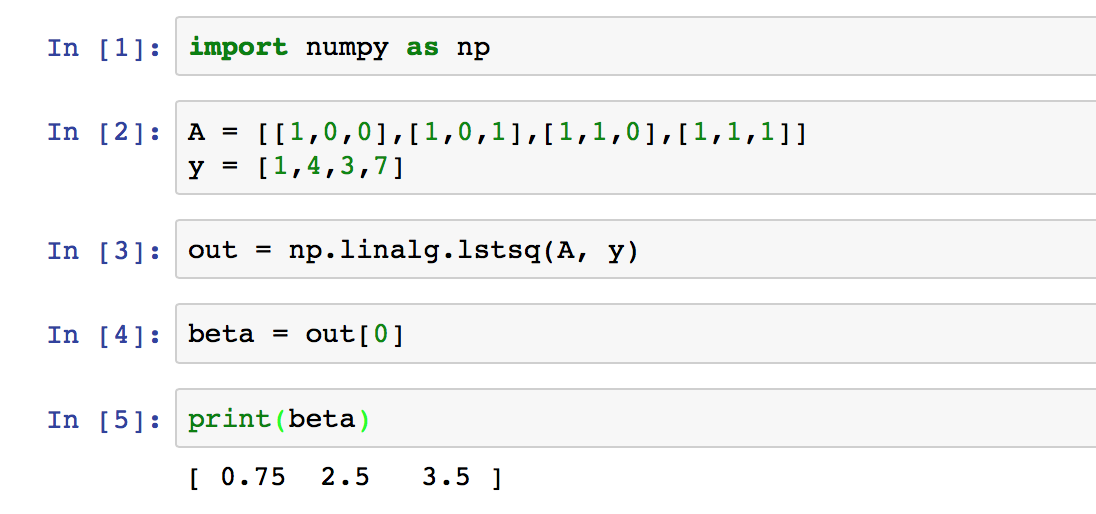
Question 1:

1. Target variable: total evaluation comes from reviews
2. , where indicates the numeric score and indicates the frequency of occurrence of words that convey judgment like “bad”, “good”, and “doesn’t work.”
3. I would change to a fraction (=numeric score/total score), instead of a numeric score.
4. , where indicates the percentage of the score, represents a rating that is simply good or bad(ex. good=1; bad=0), comes from the frequency of occurrence of some judgmental word like “bad”, “good”, and “doesn’t work”.
5. I would use (b) as a predictor. From (b), we can see the good review rate and therefore know how customers like one product. It is meaningful for us to know the total number of reviews with the word “good”.

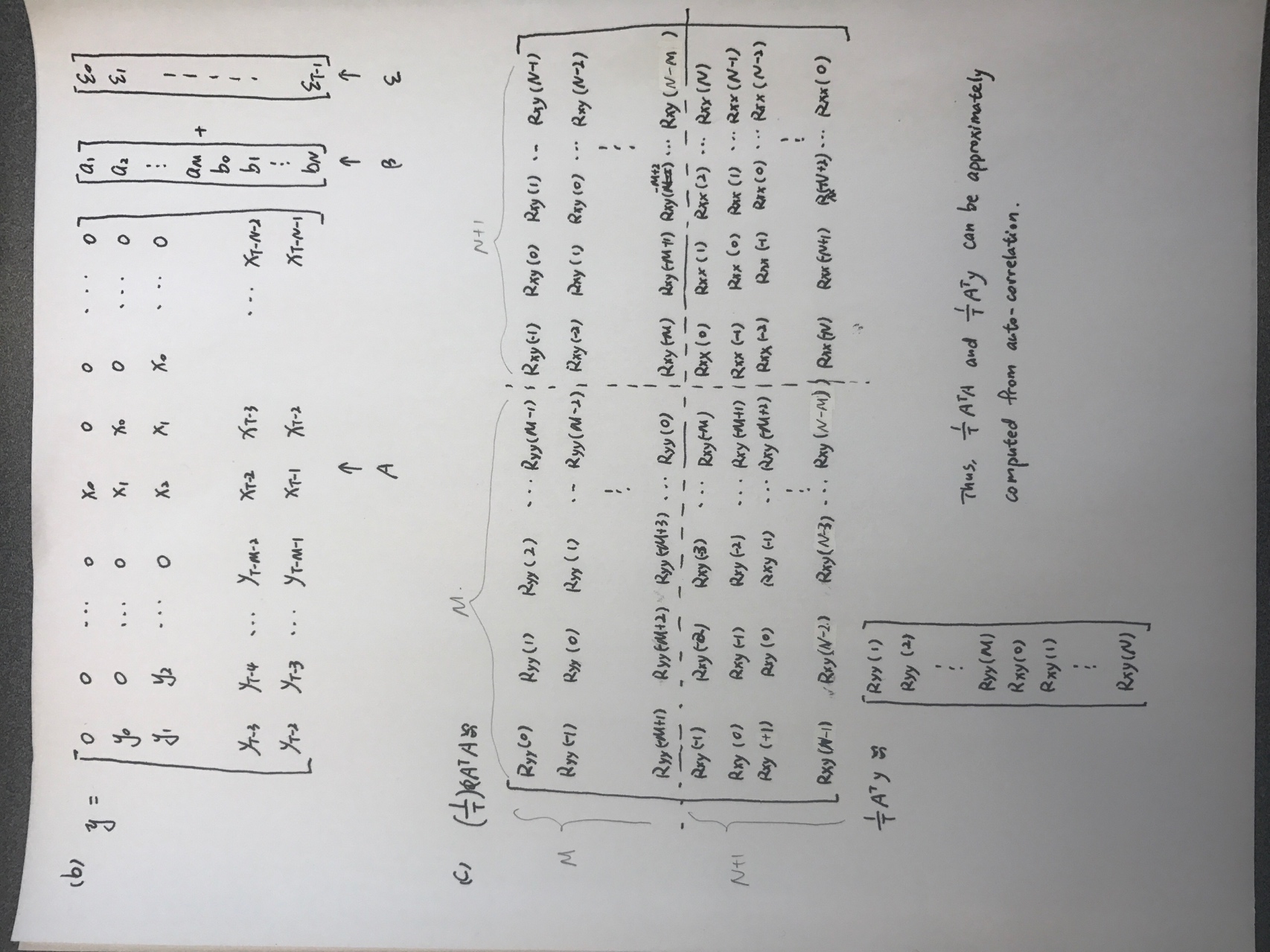
Question 2:

1. ,
2. least square solution:

Minimum RSS = 0.25



Question 3:

There are M+N+1 unknown parameters.

Question 4:

