1. The target value should be the past sales.
2. 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.
3. 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.
4. 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.
5. I would choose (b) since the total number of reviews vary, which would make result using (a) inaccurate.
6. ,
7. ,
8. ,

There are unknown parameters.

1. assume .
2. Let , .

Then can be represented as .

To solve and , we have equations. There are and , thus we can find them by solving equations in if .

1. No. if becomes unknown, will be a function of , which is not linear.

**图标

中度可信度描述已自动生成**

**文本

描述已自动生成**

**文本

低可信度描述已自动生成**