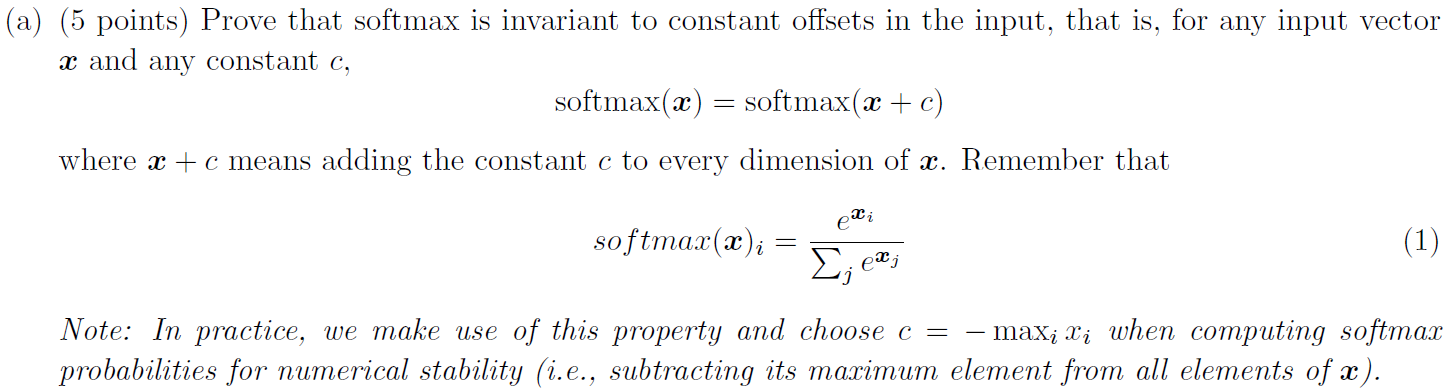
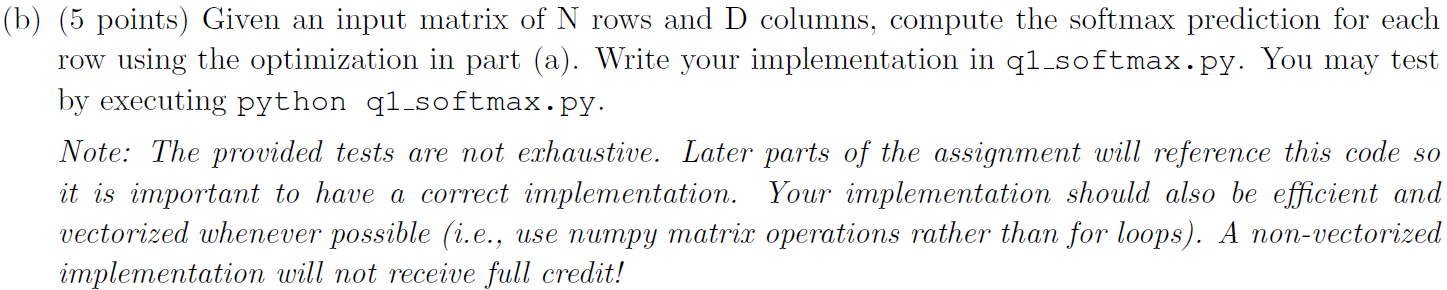
## Softmax (10 points)

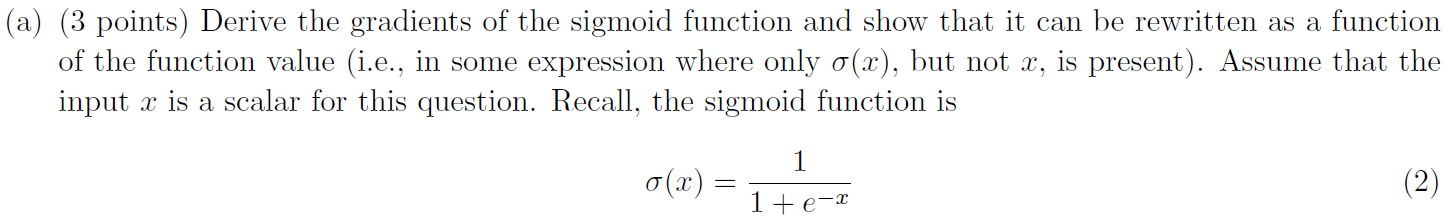


证明：



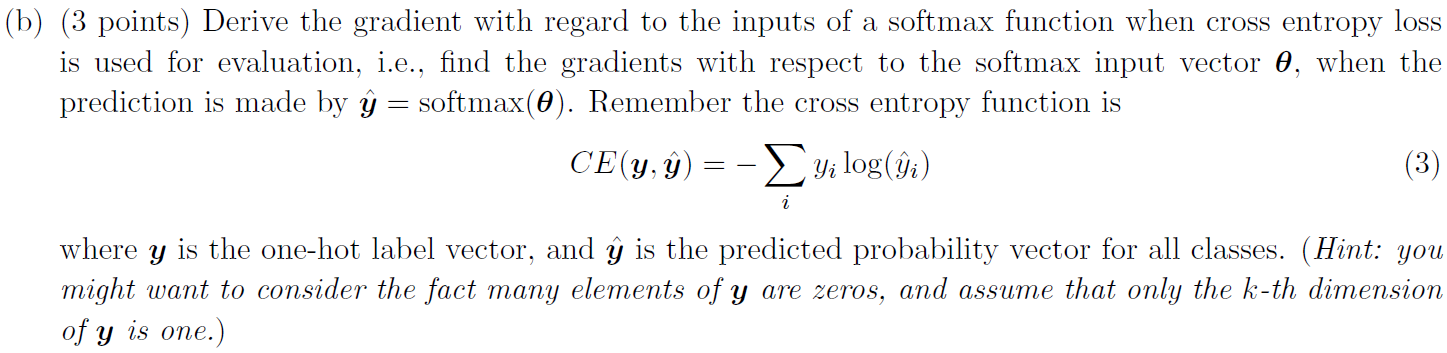
见程序。

## 2. Neural Network Basics (30 points)



解：





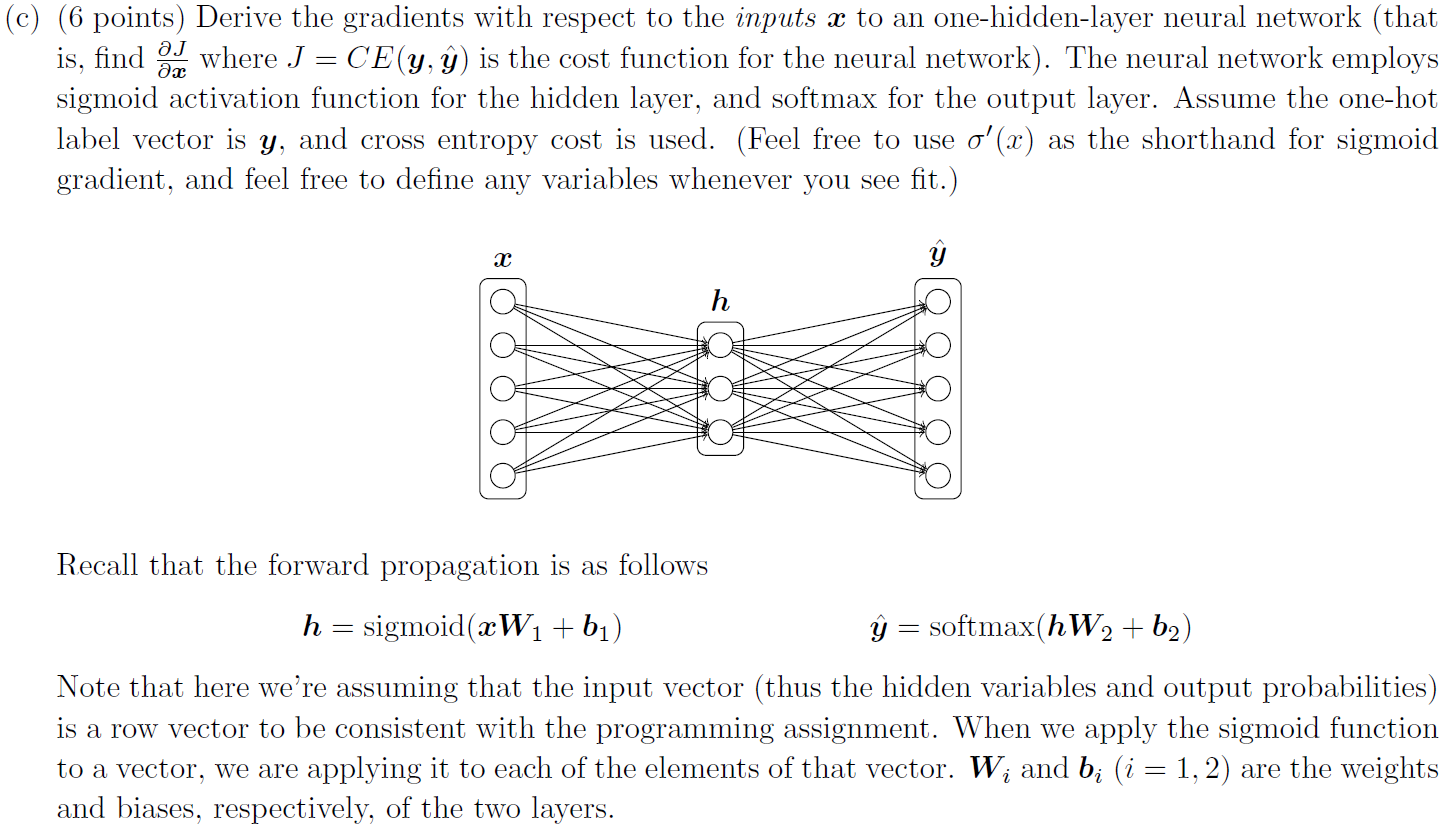
解：

，其中。k是正确类别对应的下标。

所以，实际上。







解：前向传播：



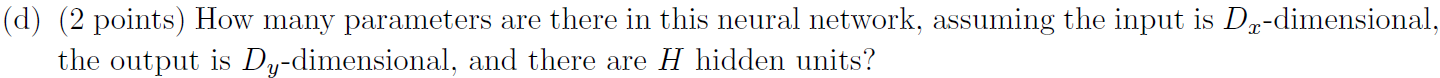
反向传播：





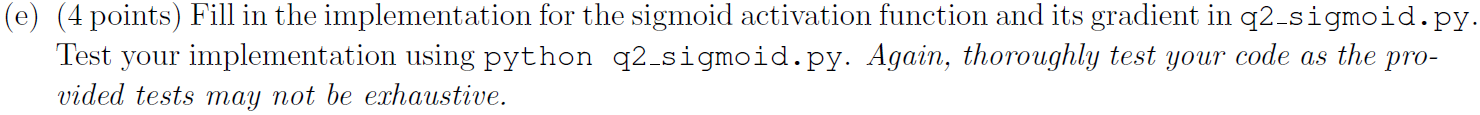




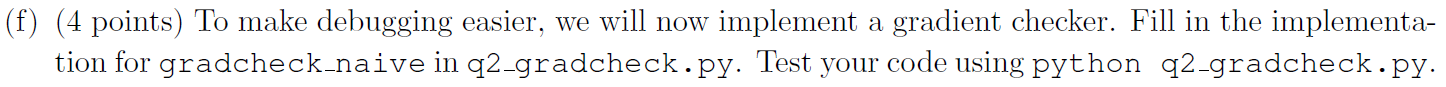


解：第1个仿射层：；第二个仿射层：。

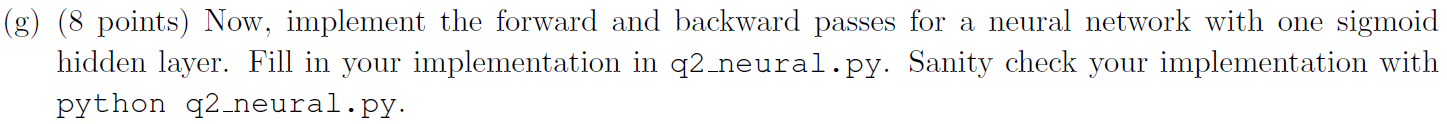
总共的参数个数：。



见程序。

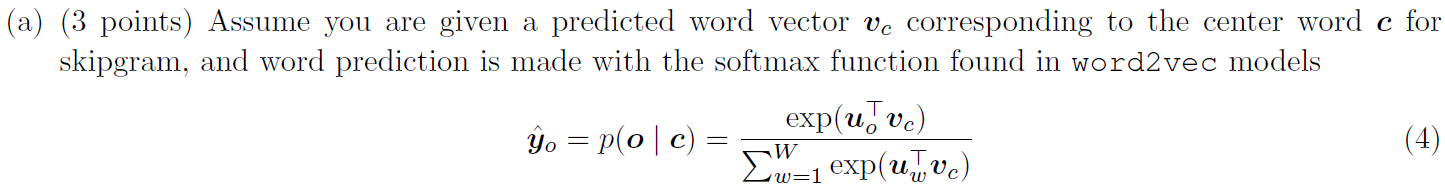


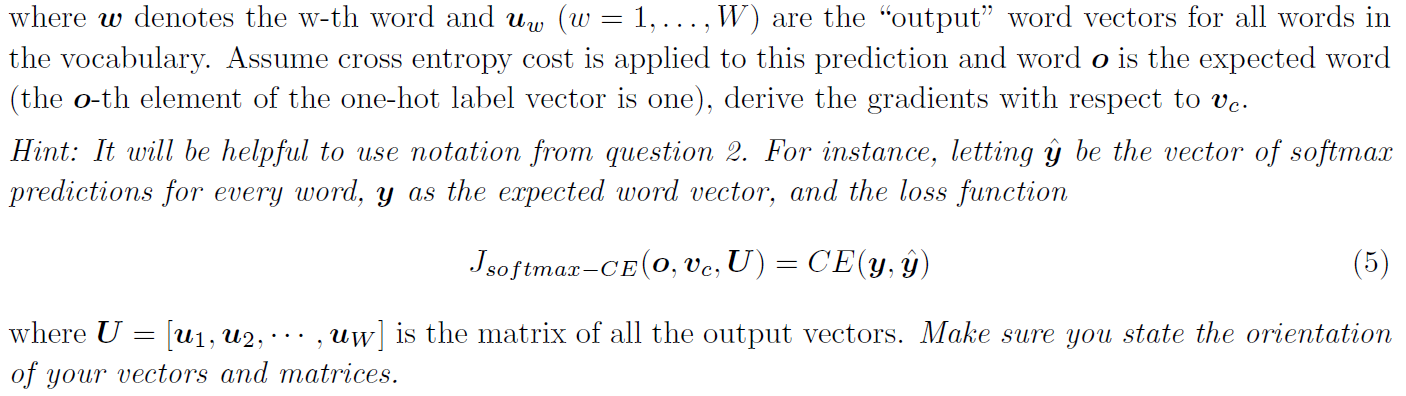
见程序。



见程序。

## 3. word2vec (40 points + 2 bonus)





解：





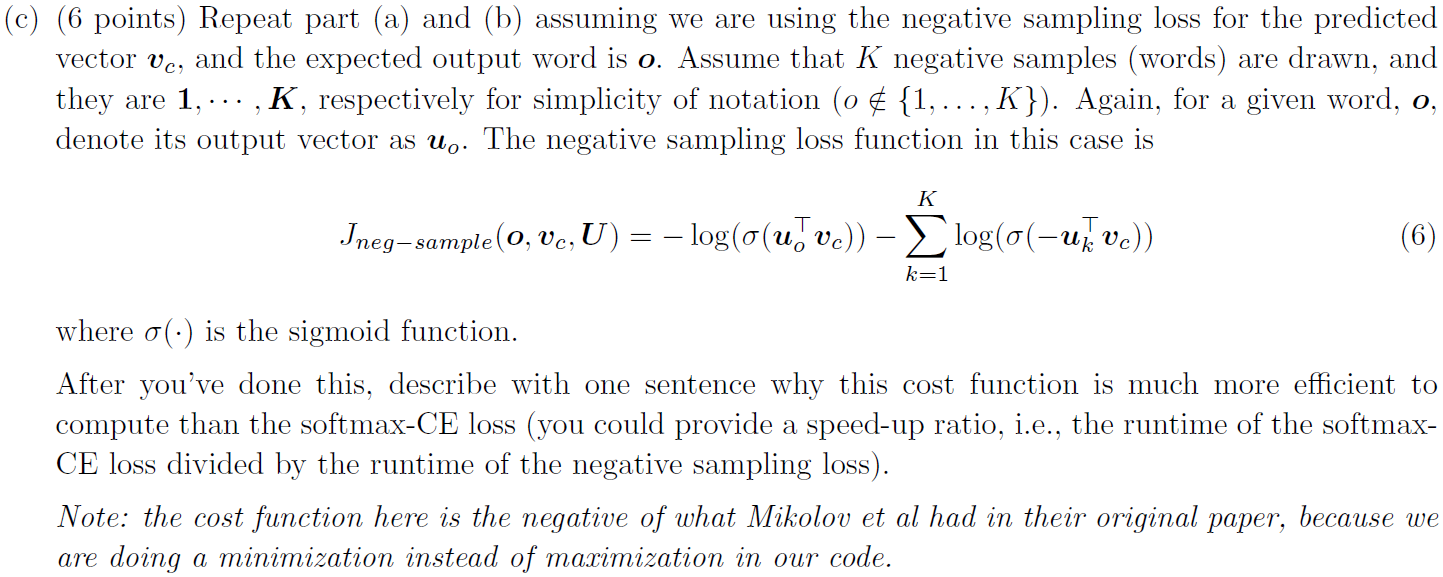


由(a)知，

当时，

当时，

所以，也即



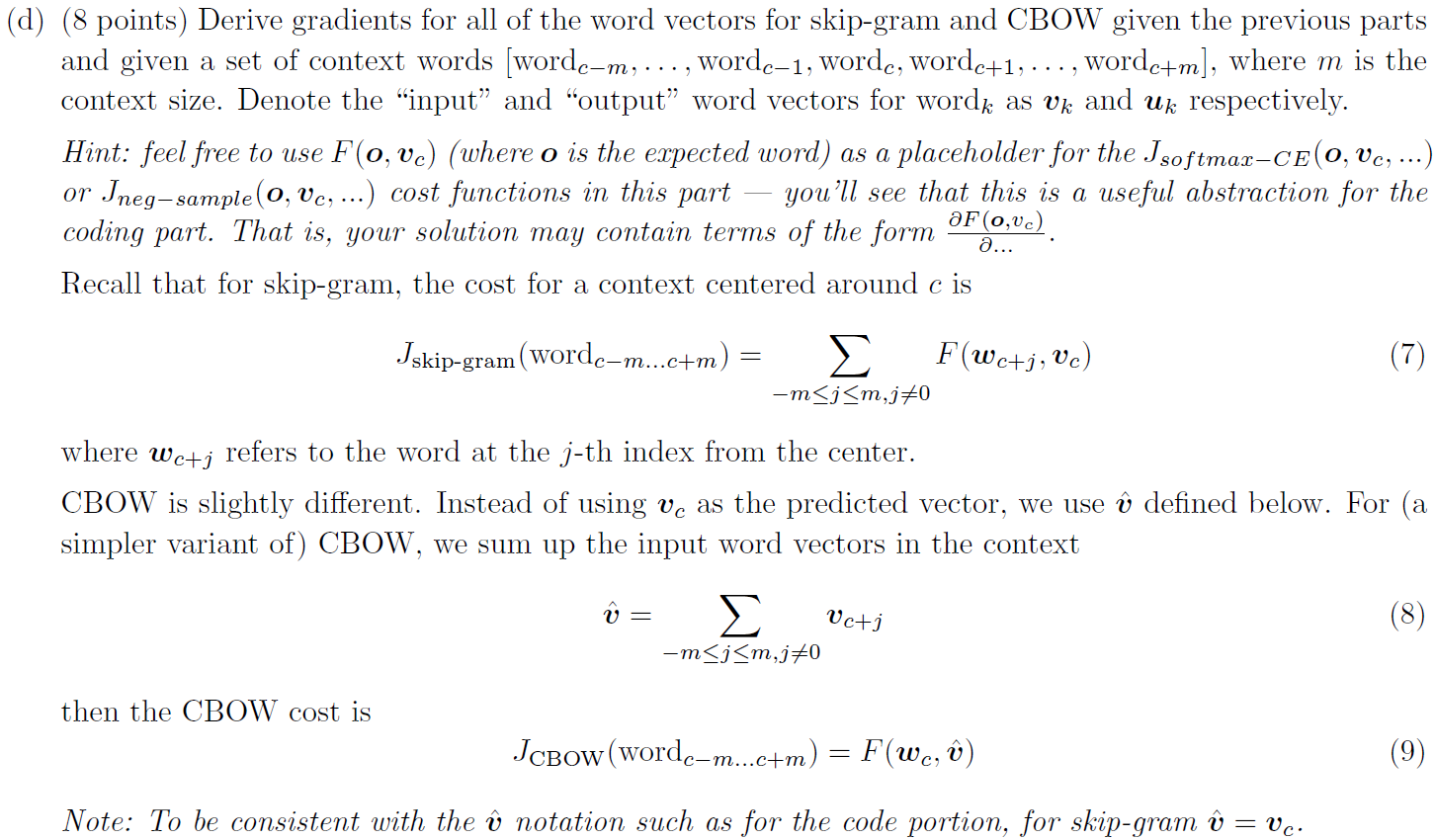
解：



需要注意的是, 所以对求导时，不能考虑的情况。







解：U是词汇集中所有单词的词向量的集合。对于给定的损失函数F，由(a)、(b)、(c)可得和。

因此，对于Skip-Gram模型，上下文窗口的损失函数的梯度如下：





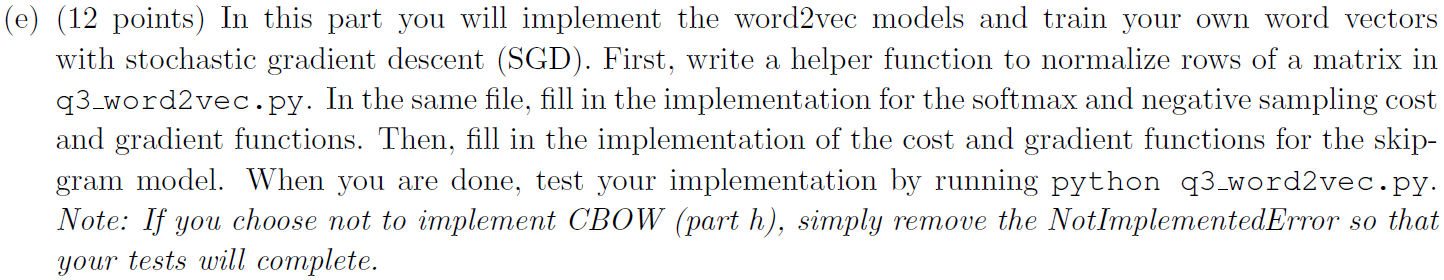


对于CBOW模型，上下文窗口的损失函数的梯度如下：

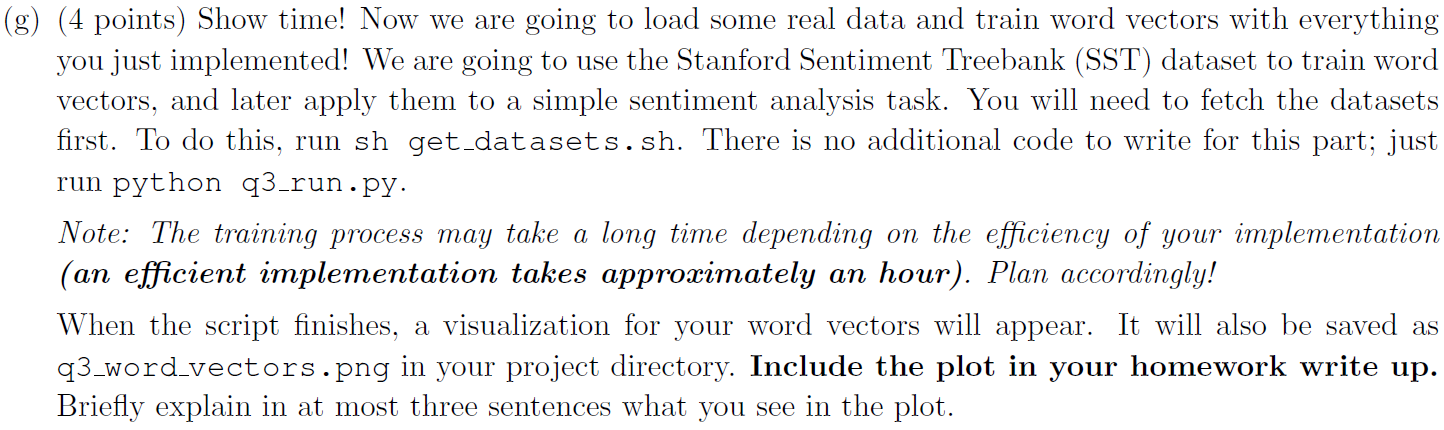


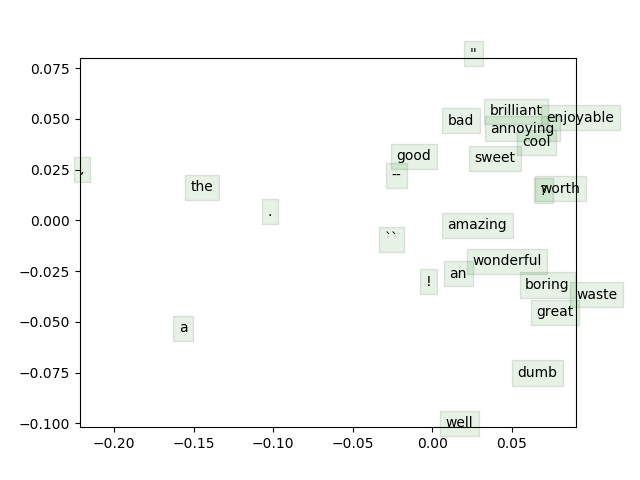




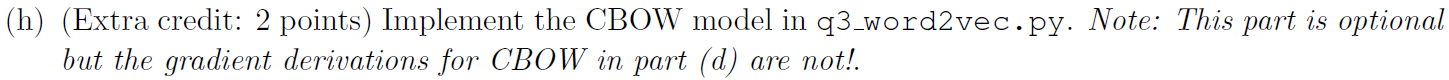








语意相似或相反的词分布在相近的区域，主要原因是因为它们的上下文比较相似，Word2Vec模型很好地捕捉到了这种相似性。



## 4. Sentiment Analysis (20 points)

