大作业二：实现Word2vec

截止时间 2021 12月 13， 23:59 之前 **得分** 15 **提交** 一份上传文件 **文件类型** zip
可用 2021 11月 22 15:40 至 2021 12月 13 23:59 21 天

此作业锁定于 2021 12月 13 23:59。

本作业需要大家实现并训练word2vec的CBOW算法。

环境需求：

- Python>=3.6
- numpy
- scipy

详细描述：

本作业需要大家实现并训练word2vec的CBOW算法（softmax版本），项目代码已上传到**此处**，请补全word2vec.py中的train_one_step函数，并成功运行python main.py。

main.py中包含3个测试点：

- test1: 用于调试模型，如果实现正确，那么最终的loss将会停留在1.0左右，且'i','he','she'三者的相似性较高。
- test2: 用实现的模型在data/treebank.txt上训练10个epoch。此部分最终的loss将会降至7.0左右，耗时约1.5h，请合理安排训练时间。
- test3: 用test2训练的模型测试效果，如果spearman相关系数高于0.3且pearson相关系数高于0.4，则通过测试。

作业提交要求：

1. PDF格式的报告中包含关键代码实现和程序的重要输出结果。
2. 整个项目代码。如有其他特殊说明，请在报告中写明。
3. 请把报告和代码打包成zip文件上传。

其他事项

1. 关于word2vec训练方式，可以参考论文：**word2vec Parameter Learning Explained (http://arxiv.org/abs/1411.2738)** 。
2. 对于机器学习中如反向传播等概念了解不足的同学，可以参考b站上吴恩达老师的课程 **https://www.bilibili.com/video/BV164411b7dx?from=search&seid=5390781987883191533&spm_id_from=333.337.0.0**

(https://www.bilibili.com/video/BV164411b7dx?from=search&seid=5390781987883191533&spm_id_from=333.337.0.0) (http://arxiv.org/abs/1411.2738)

3. 优先个人独立完成。在个人完成困难的情况下，可选择组队实现（至多2人），但需要在报告中注明各自贡献。