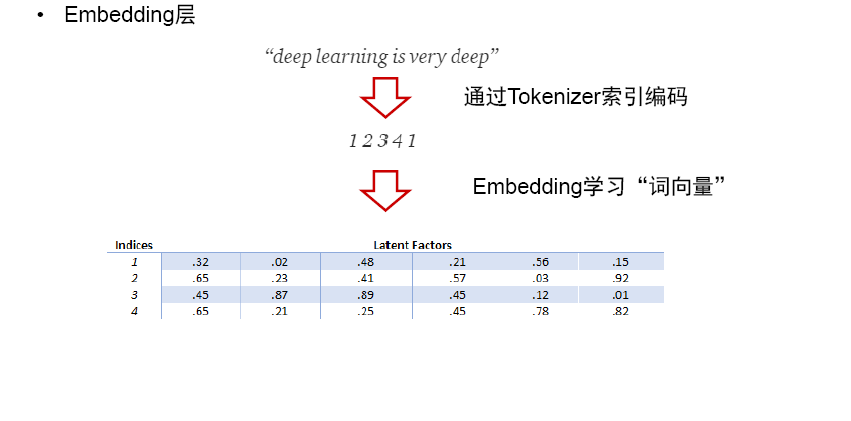
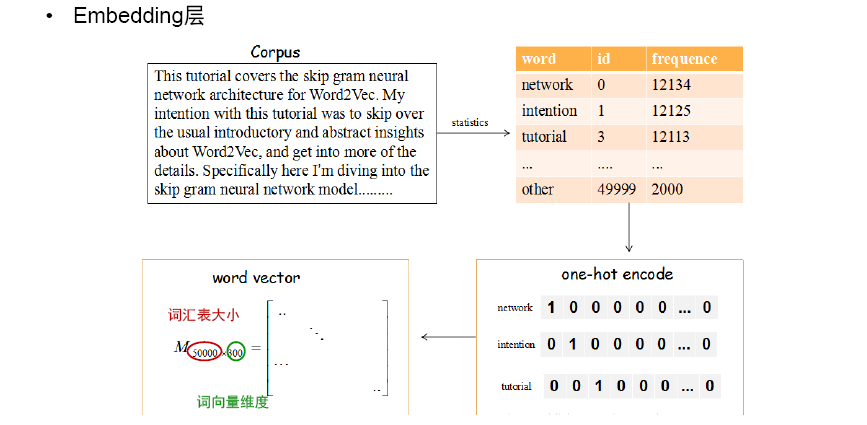
1. Twiterr 情感数据分析

利用给的数据集完成情感分析。

提示：





* 数据加载时，由于数据集中某些记录不规整，比如：包含了5列数据。所以在使用pandas进行读取时，使用了error\_bad\_lines=False 。另外可以通过调整config.n\_samples 的值控制读取的记录条数。
* 文本加载完后，使用Tokenizer 对文本进行了分词、索引编码操作，构造词汇表。keras的Tokenizer 已经包含了去除特殊字符的操作，所以这里没有做多多的处理。学员可考虑进一步处理文本数据。最后使用索引编码对文本进行处理，即tokenizer.texts\_to\_sequences(raw\_text\_list)
* 搭建模型时，第一层使用了Embedding 层对文本数据进行编码，之后使用LSTM 进行搭建。
* 模型训练过程中，使用ModelCheckPoint 对中间过程的网络参数进行保存，这里按照模型在验证集上的性能，只保存最优的训练参数。在fit 方法中，需要为callbacks 参数提供回调函数完成该功能。
* 预测时，可以加载之前保存的最优模型参数。注意，这里保存的是模型的参数，不是模型。所以，在预测时，仍然需要调用build\_model() 获取网络架构，然后使用load\_weights 加载模型参数，从而完成预测。

2.Seq to Seq 模型学习和使用

<https://blog.csdn.net/liuchonge/article/details/78824572>

自选一个数据集，完成seq to seq 模型的实现过程。

作业需要包括报告、主要环境说明、和主要代码