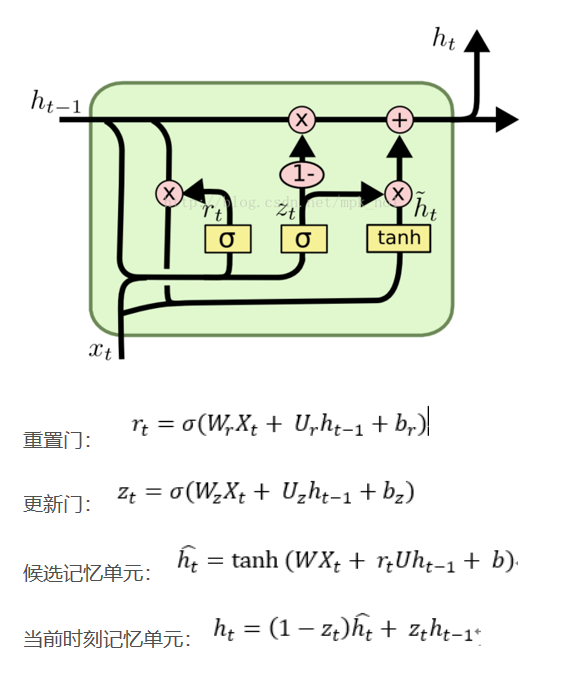
1.生成txt文件包含序列，label和序列id（用来对应pssm特征文件）。

2.学习onehot编码，把输入序列转为onehot编码

3.研究模型：2cnn+双向gru

Embedding层输入800\*20的单条序列onehot编码，输出自定义量\*800（调参），再和pssm特征做连接，得到800\*（参数+20）维作为cnn（学习序列中较浅的连接）输入。输入卷积层，选择卷积核的大小和过滤器的个数及步长，接dropout层防止过拟合，再重复上述操作，然后接池化层（最大池化），定义池化窗口和步长（减小维度），接batchnormalization层，将前一层的激活

值重新规范化，即使得其输出数据的均值接近0，其标准差接近1，接下来全连接两层Bigru（学习序列中较深的连接），双向gru是由两个gru上下叠加在一起组成的。输出由这两个gru的状态共同决定，当前值更新过程：



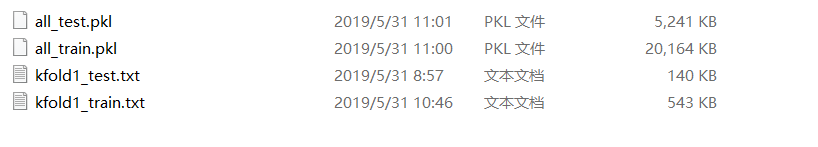
自定义输出维度（需调参），接flatten层，把多维的输入一维化，最后接全连接层和输出层（自定义输出节点个数）。

4.学习五折交叉验证，把全集分为5份，每次选出其中一份作为测试集，4份作为训练集。

方法：把每一个分类分成五份写入测试集，把全集减去测试集得到训练集。



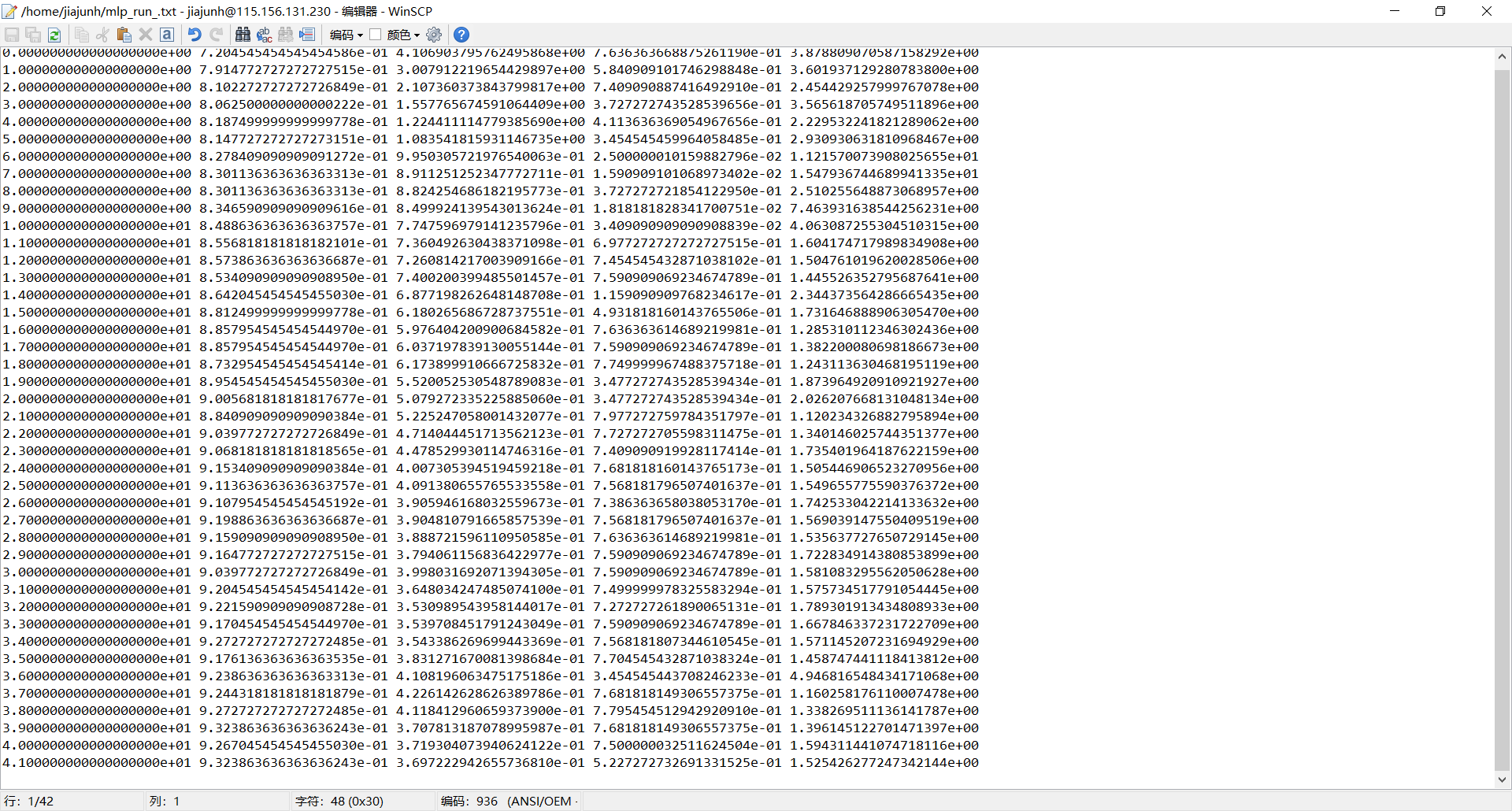
每一个文档中：



5.生成测试训练集的pickle文件

6.由于gpu被占用，周二改完模型未能提交，周四换节点重配环境跑了一折周五服务器崩了..，一折的结果：

轮数，训练集acc，训练集loss，测试集acc，测试集loss：



7.接下来的任务：跑完五折交叉验证查看mean（score），调参，寻找独立测试集，比较该模型与其他模型的优劣，查看是否需要修改模型。