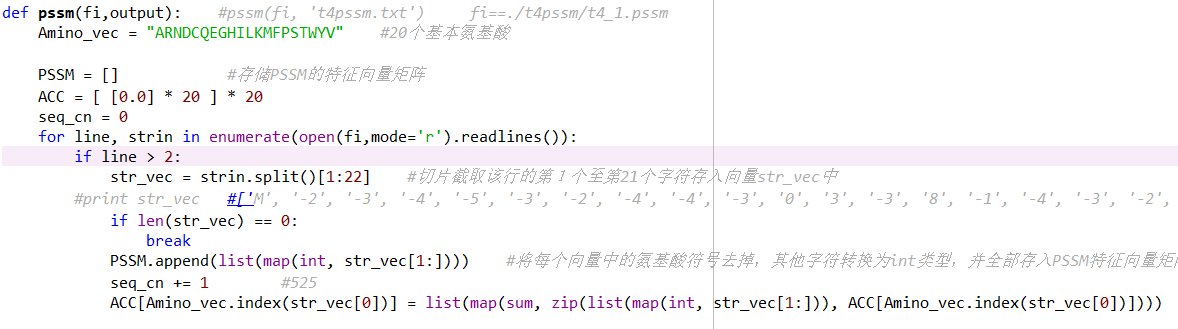
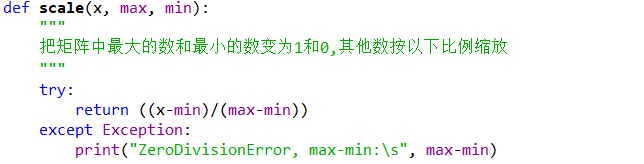
# PSSM的处理

由于刚刚注册自己的服务器账号，下载WINSCP并大致了解了一下功能，由于把数据从王雷学长那里直接传送耗时太久，所以建立软链接，使用在他服务器账号下跑好的氧结合蛋白的数据。

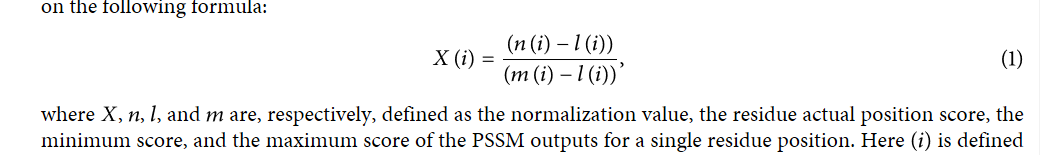
大致思路：读取pssm文件，提取20\*20的矩阵，进行归一化，再打成400\*1的向量。



把第一第二行跳过，从第三行开始提取，和第2到21列的数据提取，同时计数，得到氨基酸链的长度.



归一化过程，根据论文中的公式



进行归一化



把20\*20的矩阵用reshape函数将矩阵的行列维度重新调整为400\*1，concatenate进行连接操作，合成新矩阵。

测试结果：



对单个的处理完成了，但是对所有的pssm的文件特征的批量提取还未完成，在服务器上跑的时候出现报错，还未解决。