基础篇：

1. 如何使用远程连接从windows或者linux连到服务器进行操作(切换用户，传递文件) (严禁使用root账户)
2. Linux基本bash命令
   1. 查看文件大小，内存大小，cpu信息，硬盘空间等等
   2. 查看目前运行程序情况，剩余内存，kill程序
   3. 运行一个程序
   4. 写一个脚本，使用上for, if等语句
   5. sed命令, grep命令
3. Awk练习
   1. 词、字的统计
   2. 取出某些特定的字段
   3. 改写：根据某个特定字段进行排序
4. 程序运行实践
   1. Stanford nlp: 分词、词性标注、句法分析； 整体跑一遍
   2. HIT-ltp: 分词、词性标注、句法分析； 整体跑一遍
   3. CMake使用
   4. ZPar: 分词、词性标注、句法分析； 整体跑一遍（模型自己训练）
   5. Berkeley parser使用 （模型自己训练）
   6. 情感分析代码跑一遍：NNContextSentiment
   7. NNCRF跑一遍：NNSegmentation, NNPostagging, NNNameEntity
   8. 使用word2vec训练一个词向量
5. 使用github，建立自己的github，建立自己的代码和文档仓库
6. 查找一个你认为比较好用的开源应用软件(大众熟知的机器学习或者自然语言处理工具)：例如我推荐 beyond compare, ConstTreeViewer, whatswrong
7. 致力于研究的同学，使用latex(CTex工具包)，使用yinxiang笔记

编程篇：C++必须， 然后从java和python任选一个

1. 情感分类评价程序
2. 词性标注评价程序
3. 分词评价程序
4. 命名实体识别评价程序
5. 8、9、10、11用CMake组织起来
6. 矩阵向量运算 y = Wx, (稀疏情况和dense情况)
7. Dense情况： y= W1x1 + W2x2
8. 矩阵向量运算： y = tanh(Wx), y = sigmoid(Wx) , y = relu(Wx) (稀疏情况和dense情况)
9. Dense情况： y = tanh(W1x1+W2x2), y = sigmoid(W1x1+W2x2) , y = relu(W1x1+W2x2)
10. 向量运算softmax: y = softmax(o)
11. 根据template抽取情感分析特征
12. 根据template抽取词性标注特征
13. 根据template抽取分词特征
14. 根据template抽取命名实体识别特征
15. 使用svm进行18、19、20、21
16. 利用自己抽取的特征使用NNSegmentation, NNPostagging, NNNameEntity
17. 自己动手写最大熵程序进行情感分析，使用template抽取的特征，学习什么是最大熵
18. 自己动手写最大熵程序进行情感分析，使用外部word2vec特征，学习什么是pooling
19. 自己动手写最大熵程序进行情感分析，词向量本身作为参数模型，第一个深度学习模型
20. 24和25进行combine后预测
21. 使用max-margin重做24、25、26、27
22. 加入额外普通隐层进行25、26、27
23. 加入额外RNN隐层进行25、26、27
24. 加入额外LSTM隐层进行25、26、27
25. 28基础上将pooling换成attention进行25、26、27
26. 29基础上，将character向量引入，进行25、26、27