**回测中择时的机器学习模型**

韩昊阳

本周主要的工作内容为：

1. 配置36服务器：

配置的主要部分有：

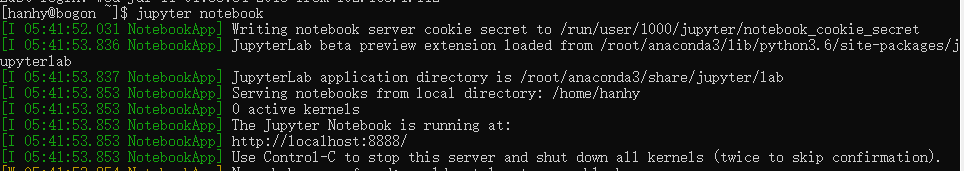
1. Anaconda 3 with python 3.6
2. Cuda 9.0
3. Cudnn 7.1.2(for pytorch)
4. Tensorflow
5. Keras( generally tensorflow backened)
6. Pytorch
7. Theano
8. 安装了pip conda等一系列安装包的terminal 命令

经过测试，我们可以在本地电脑中可视化的使用jupyter notebook了，具体方法为（以36服务器我的账号hanhy(password:123456)为例）：

*安装ubuntu for windows10(或者wolf等terminal软件)；*

*输入 ssh* [*hanhy@192.168.1.36*](mailto:hanhy@192.168.1.36)*；*

*进入36服务器后，输入jupyter notebook，等待返回的url地址，并记录端口（图示为8888）（第一次使用新账号输入jupyter notebook会在url后返回token）；*

**

*随后打开新的bash界面，输入 ssh -L 8000:localhost:8888* [*hanhy@192.168.1.36*](mailto:hanhy@192.168.1.36)*, 其中8000是本地ip，覆盖了36服务器中的ip:8888；*

*在浏览器中输入localhost: 8000 进入jupyter notebook，第一次使用会要求输入token或者设置密码。*

*注意：如果使用root账号登录jupyter notebook需要使用：jupyter notebook –allow-root命令。*

*注意：使用普通权限账号登录36可能会遇到permission deny的问题，可以换成root账号做暂时操作。*

深度学习库已经安装并调试完毕，阳雄哥也做了很多工作。需要注意的是，使用tensorflow跑程序（包括tensorflow backened的keras跑程序）会占用90%以上的gpu计算力（总共8000M左右，tensorflow会占用7400M+），此时还可以跑一个pytorch程序（pytorch是使用gpu可调整的）。

在日后的安装过程中应当注意版本匹配的问题（tensorflow 1.8+cuda 9.0 + cudnn 7.0）。如果要使用theano应修改keras的参数设置。

1. 标准化机器学习流程。

本周将前几周的工作标准化，生成了针对回测中选股、择时步骤的完整流程python测试库。将其附在最后，大约分为：

1. Importer
2. Normalizer
3. Splitter
4. Fitter
5. Evaluator
6. Total Process

六个部分。其中针对选股和择时的函数有不同差异。将这个版本的测试库附在最后。

1. 对回测中择时的主流机器学习模型进行benchmark测试（使用与选股相同的超参数）

随后对择时模型进行基准测试，使用了evaluator中的simple\_analysis函数返回IC、IR、Sharpe值。由于未将所有因子放入bsim平台中进行综合评测，评估标准的完整程度和精确性无法保证。测试了Linear, Lasso, Ridge, Bayesian Ridge, Decision Tree, Random Forest, GBDT, Bagging, SVM regression 模型的测试，并保留了拟合后的因子，等周一可以得知拟合结果如何，并逐个对模型进行改进。