

1. 按参考资料配置 Hadoop 集群。
2. 开启集群。

```
cd /usr/local/hadoop/  
bin/hdfs namenode -format  
sbin/start-dfs.sh  
sbin/start-yarn.sh
```

成功开启集群后，在 master 的终端运行 jps 检查集群状态（不同的配置会有不同的状态）：

```
hadoop@master:/usr/local/hadoop$ jps  
3522 Jps  
2951 ResourceManager  
2797 SecondaryNameNode  
2590 NameNode
```

3. 先将 data 文件夹中的文件放在/usr/local/hadoop/code/路径下，再按下面的命令将其拷贝至 hdfs。

```
cd /usr/local/hadoop/  
sudo bin/hadoop fs -copyFromLocal code/test_samples.txt  
/user/hadoop/test_samples.txt  
sudo bin/hadoop fs -copyFromLocal code/train_samples.txt  
/user/hadoop/train_samples.txt
```

4. 先将 src 文件夹中的文件放在/usr/local/hadoop/code/路径下，再按下面的命令进行编译。要注意的是，代码实现用到了 commons-io.jar 包，该包在我的计算机上的路径是 /usr/local/Apache/commons-io-2.6/commons-io-2.6.jar，不同计算机上的 commons-io.jar 包的路径可能不同。也可以避开编译过程，直接执行编译好的 KNN 程序（第 5 步）。

```
cd /usr/local/hadoop/code/  
sudo javac -cp ../share/hadoop/common/hadoop-common-  
2.6.0.jar:../share/hadoop/mapreduce/hadoop-mapreduce-client-core-  
2.6.0.jar:../share/hadoop/common/lib/commons-cli-  
1.2.jar:/usr/local/Apache/commons-io-2.6/commons-io-2.6.jar KNN.java -  
d ./  
sudo jar -cvf KNN.jar KNN*.class
```

5. 执行程序

输入训练文件为：train_samples.txt

输入测试文件为：test_samples.txt

输出预测结果为：predictions

```
cd /usr/local/hadoop/code/  
/usr/local/hadoop/bin/hadoop jar KNN.jar KNN train_samples.txt  
predictions /user/hadoop/test_samples.txt
```

6. 观察预测结果

```
cd /usr/local/hadoop/  
bin/hdfs dfs -cat predictions/*
```

KNN 算法预测结果输出如下:

```
hadoop@master:/usr/local/hadoop$ bin/hdfs dfs -cat predictions/*  
0      6  
1      7  
2      0
```
