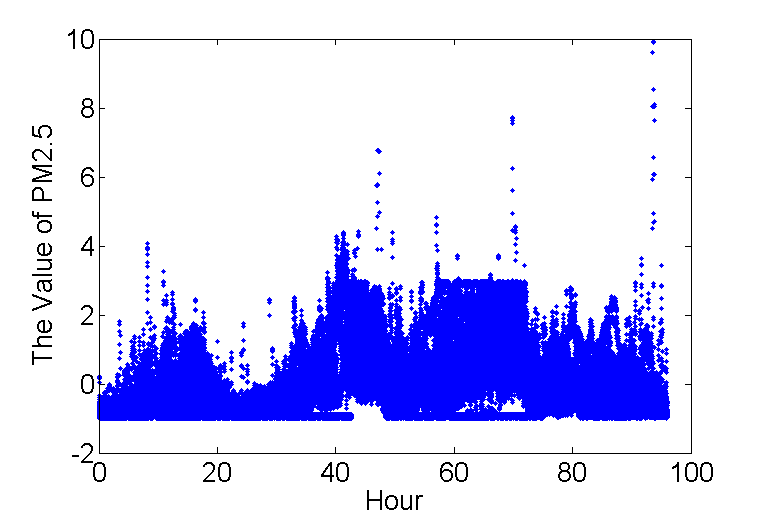
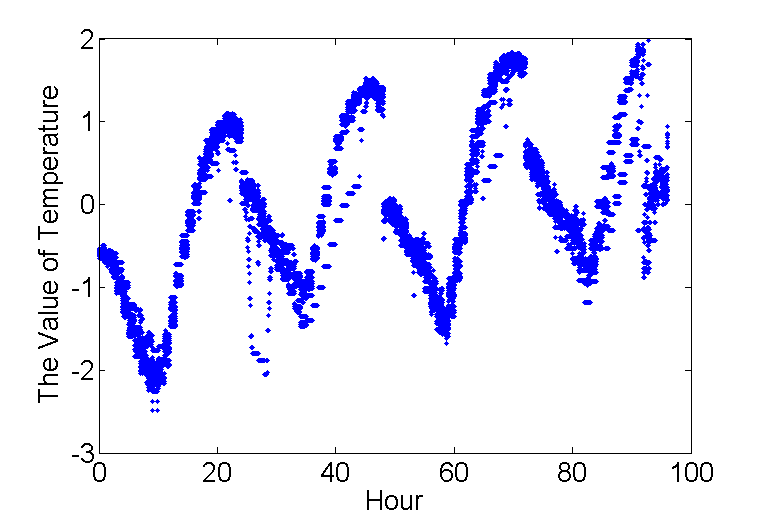
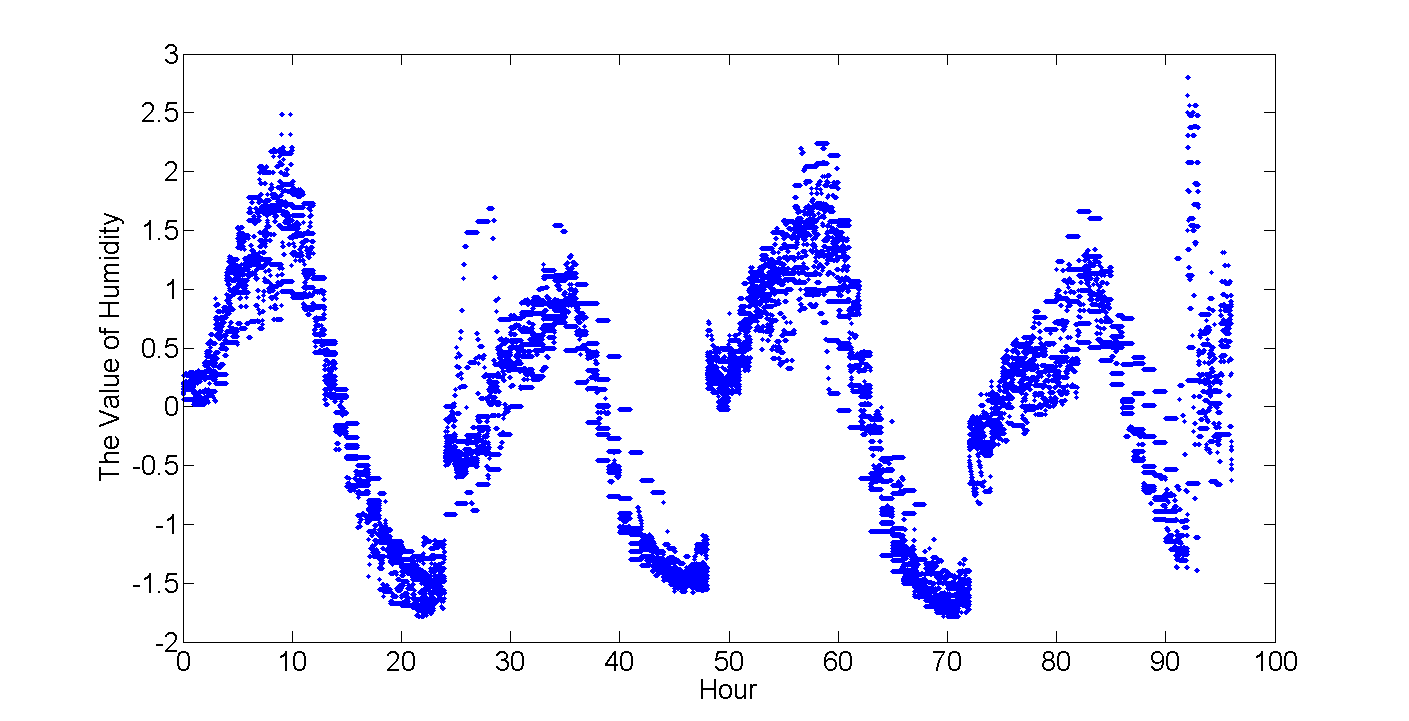
数据处理：

Z\_score方法归一化数据

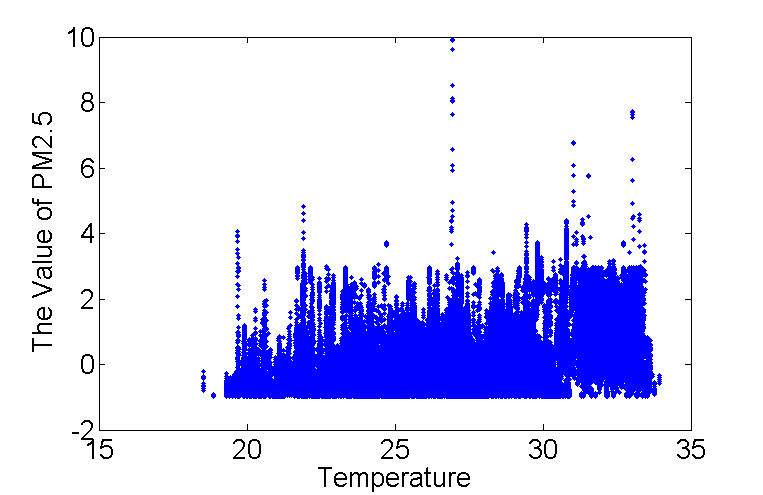
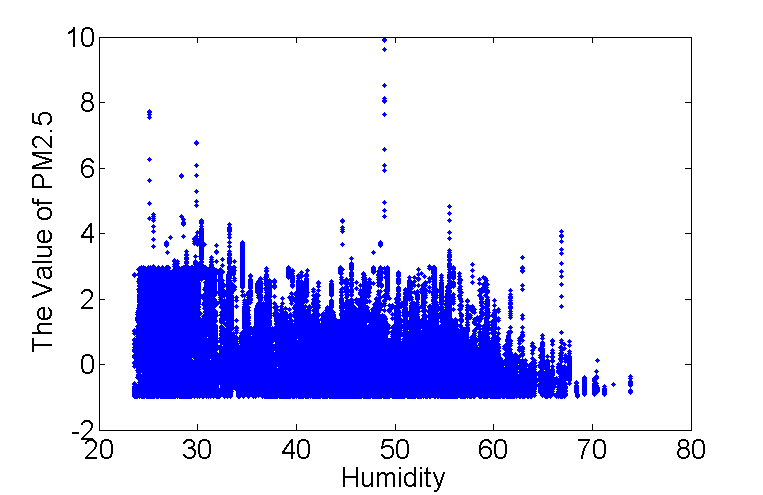
可以看出总体数据集，PM2.5在时间轴上具有周期性，4天的数据，具有四个峰值。同样压强和湿度等属性在时间轴上也具有明显的周期性。

从对比图可以看出PM2.5的噪声非常多。





而PM25在其他属性上周期性并不明显。



Lab1 : 验证 KD树一层划分结果确实保证了数据在天气和空间上的相邻相似性

通过查看划分结果的PM2.5方差来判定

从数据量5000、10000、20000、50000进行实验

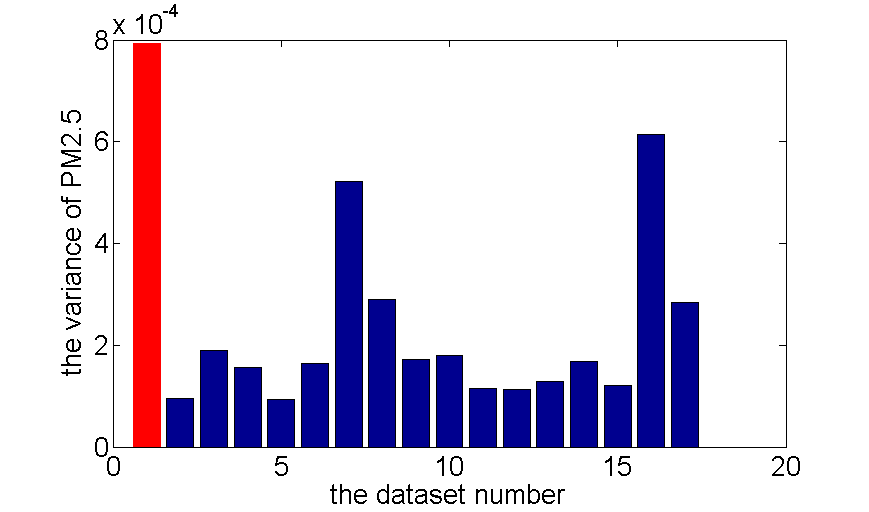
红色柱子表示划分前数据集的方差

蓝色柱子表示划分后每个划分集合的PM2.5方差

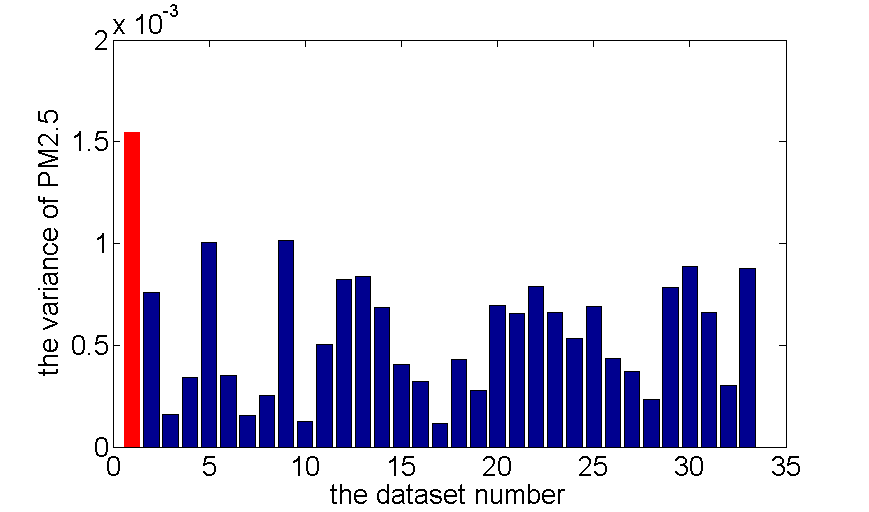
从图可以看出，划分以后数据集的PM2.5方差确实得到明显的下降

方差越小，训练误差越小

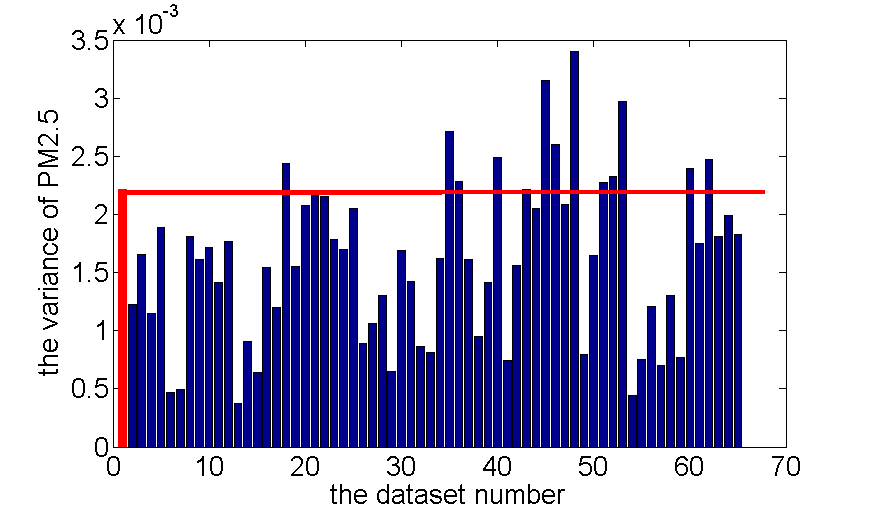
1. 5000



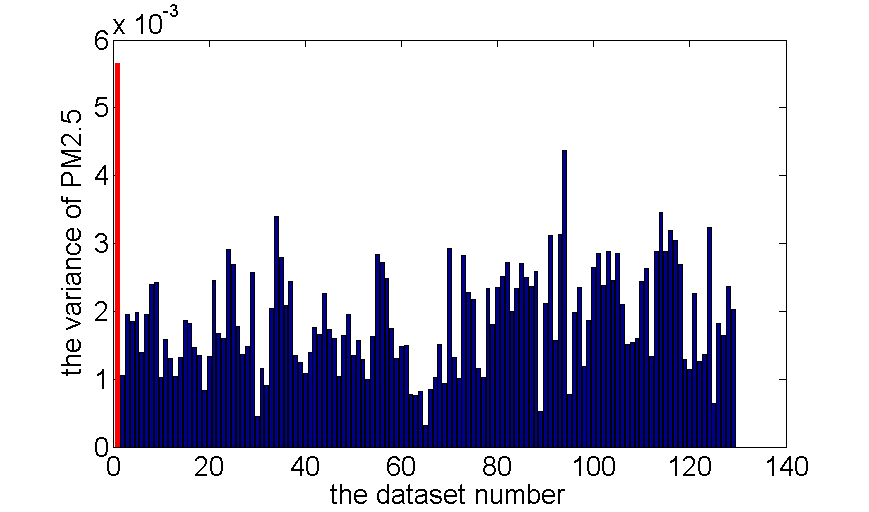
1. 10000



1. 20000



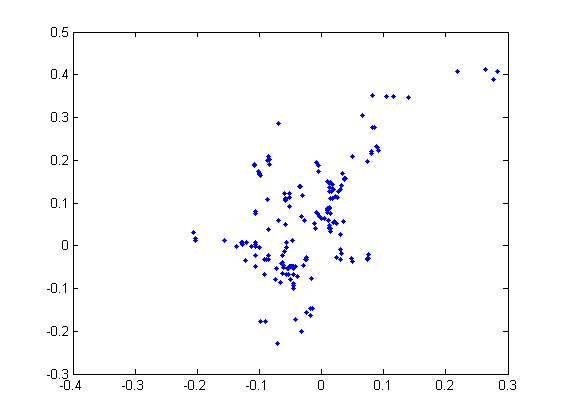
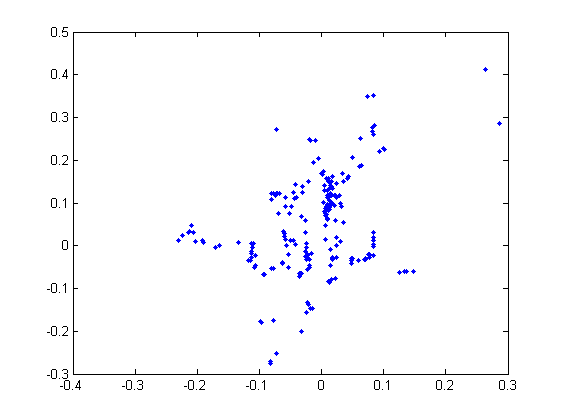
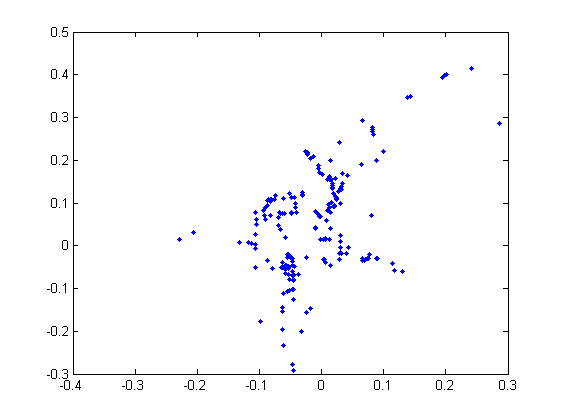
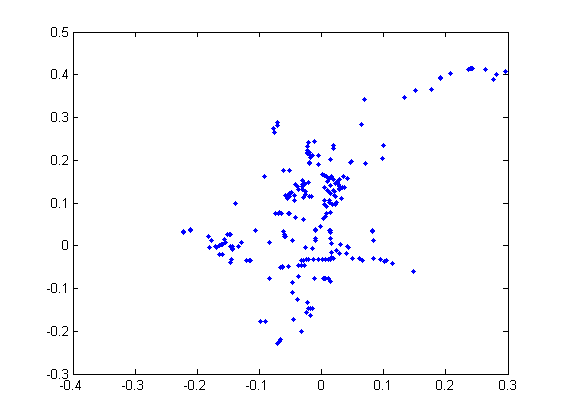
1. 50000



Lab2: 按照经纬度和天气划分得到的相邻相似数据集，在空间分布上并不完全相邻，而是散乱分布的。

所以

以下图为经纬度分布图



Lab2：

理想状况下，对数据进行一层聚类以后，就可以开始二层聚类，按照频率分析的思路，用多个正弦波去拟合子数据集（时间-PM2.5），但由于数据分布呈现如下状态：及在时间上分布不均匀，数据缺失严重，难以学出波形。

模型代码：

1. hyp\_model = initial\_model()

初始化：Q, size\_limit

1. [model, train\_time, train\_error] = train\_model (hyp\_model. trainingSet\_x, trainingSet\_y)

Model里面的参数：

流程：

tree = kd\_buildtree(trainingSet, 0);

nodeset\_cell = generateSubset(tree, 1, layer\_std); % 划分集合

预设超参数

Hyp\_subset(i) = minimize(); % 为每一个集合训练子模型

存时间、存超参数

返回

1. [test\_result, test\_time] = test\_model(model, testingSet\_x)

流程：

For 每一个测试点

Nearest\_id = findNearest() % 可能需要修改一下代码，因为已经用了7个属性划分

For 每一个自学习器

计算预测结果和权重;

End

加权和；

End

但是我们觉得我们更好的相处状态应该是，完全不要有交流负担，有聊的就多说一点，没什么聊的，那就看看对方，傻笑一会儿就可以接着干正事，我想这才是最轻松也是最有利于我们关系发展的相处方式，我觉得不是我们无话可说，而是拼搏的过程本就千篇一律，就像xiaozhe，几年的创业经历，能用每天高效工作13个小时就描述完了，后续的各中感概想必也是如今回想起来才有所出的。

我觉得只要我们都走在正确的方向上，然后和彼此相处的状态很轻松，那么我们就永远不会相距太远，因为只要一想到我们都在为共同的未来努力就觉得很温暖。我认为自己在全力以赴努力的时候，对你的爱最差也就是停滞不前，并不会减弱，因为我爱你，我想为你变得更优秀，我把你变成了自己的正能量。所以我也希望你能努力把我变成自己的正能量。