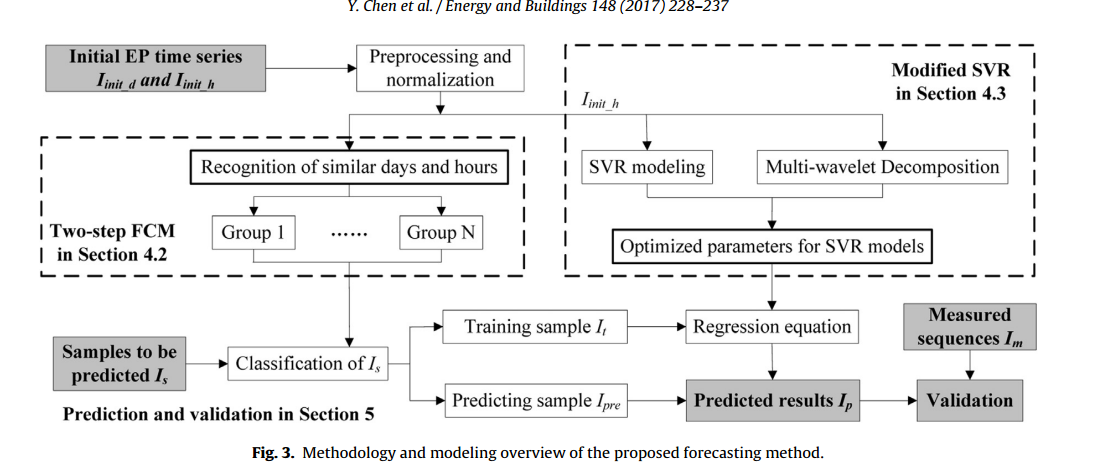
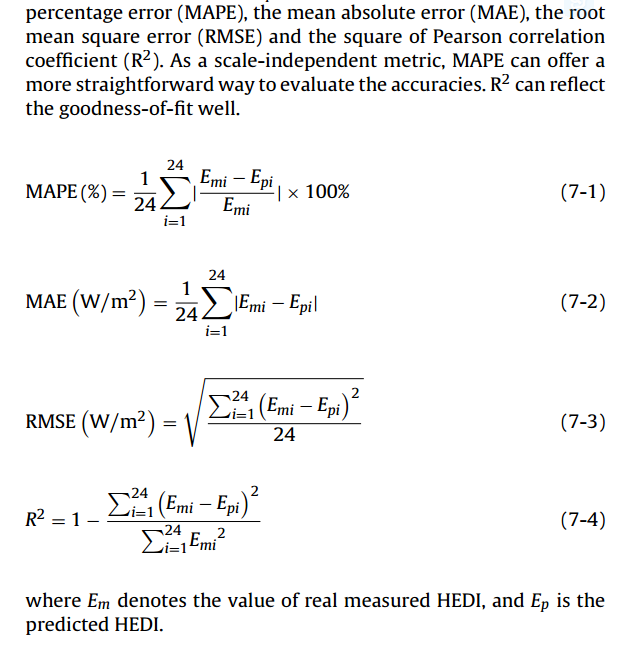
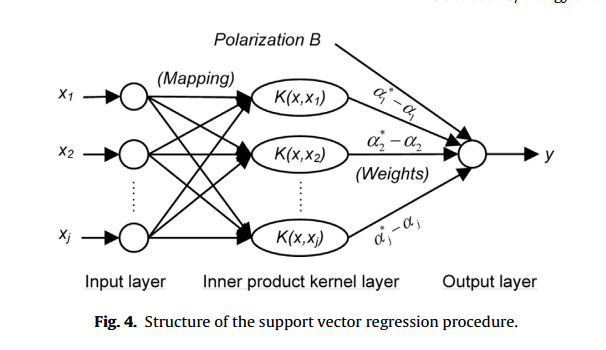
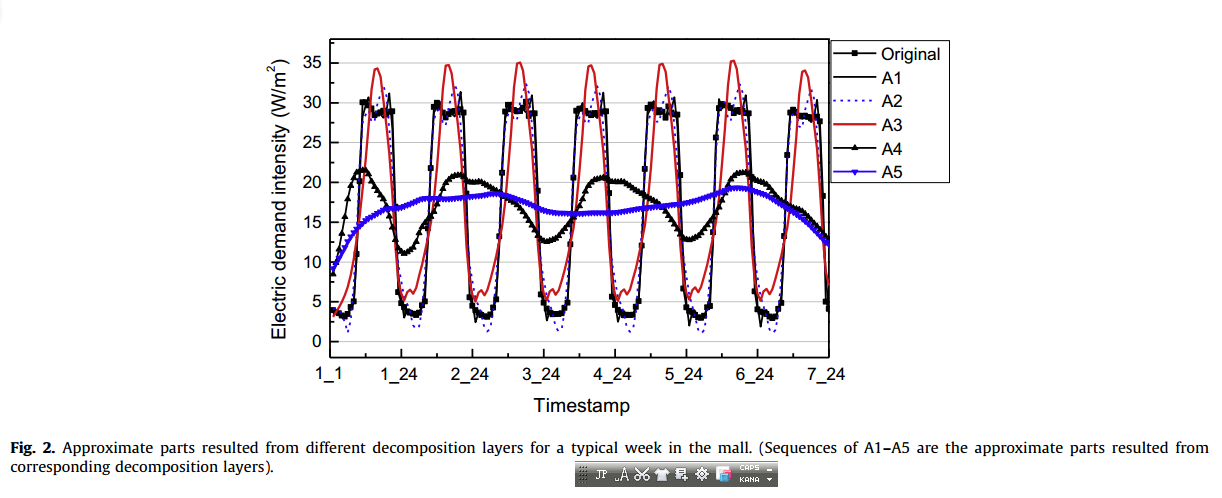
首先用FCM对酒店100天的能耗进行聚类。



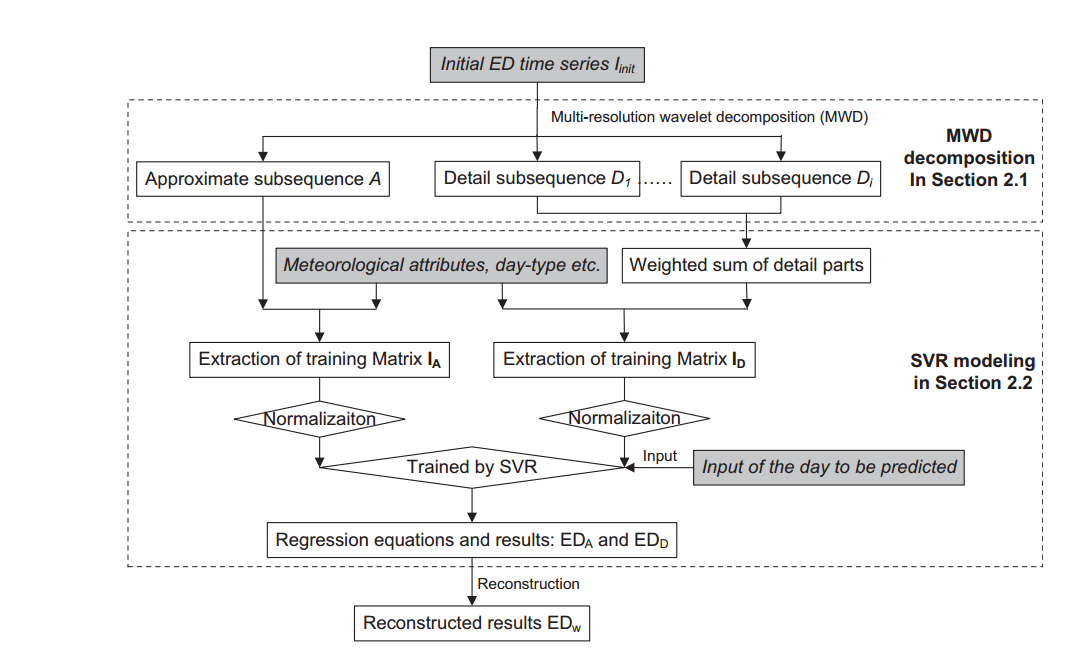
然后先用mwd（小波滤波算法，能在去除噪声的时候很好的保持信号的突变或图像的周源）来去掉误差杂，再SVR来预测能耗（分hour）。先天聚类，再对小时进行预测。



复现此算法。写成整个程序（文远楼）放弃阳光辐射。



小波分解，把能耗波分解成不同波，然后针对不同波训练再预测

我们现在要搞清楚的就是聚类之后，陈师姐是如何对聚在一起的样本天进行小波分解再svr的。

书上说 **obj[val]** 用来选取DataFrame的单个列或一组列，其方法是通过具体的columns名查询，而并不能用单纯的数字来索引：

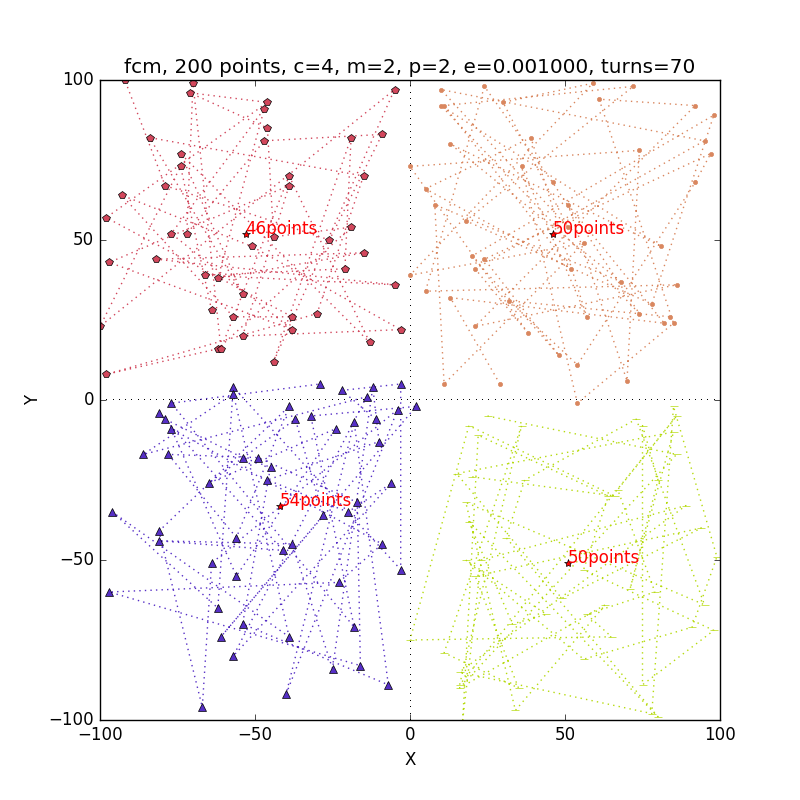
Fcm 的Python实现

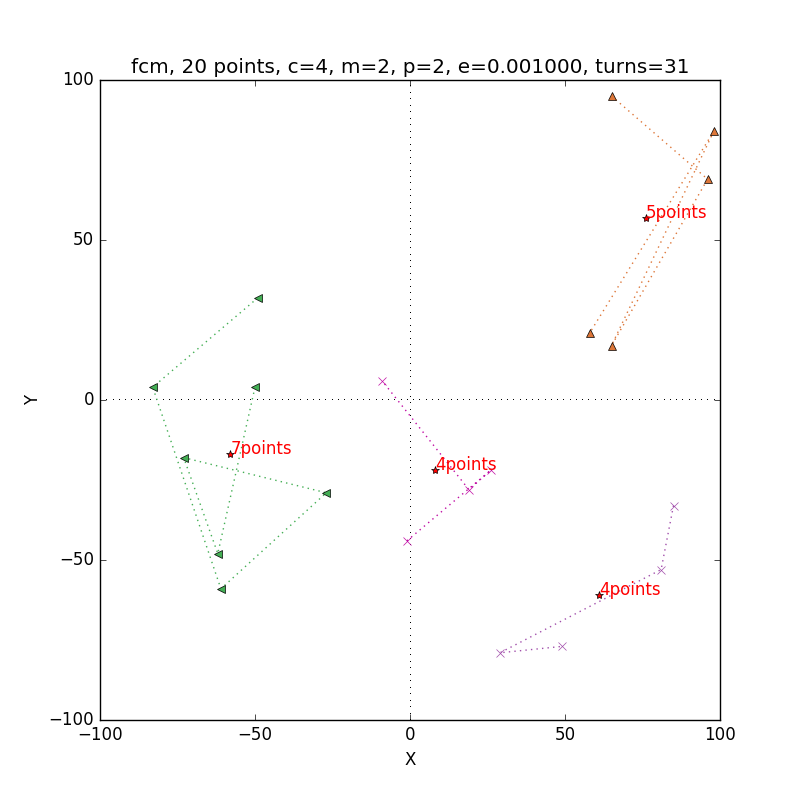
1给定同维向量数据集合points，数目为n，将其聚为C类，m为权重值，u为初始匹配度矩阵（n\*C），采用闵式距离算法，其参数为p，迭代终止条件为终止值e（取值范围(0，1））及终止轮次。

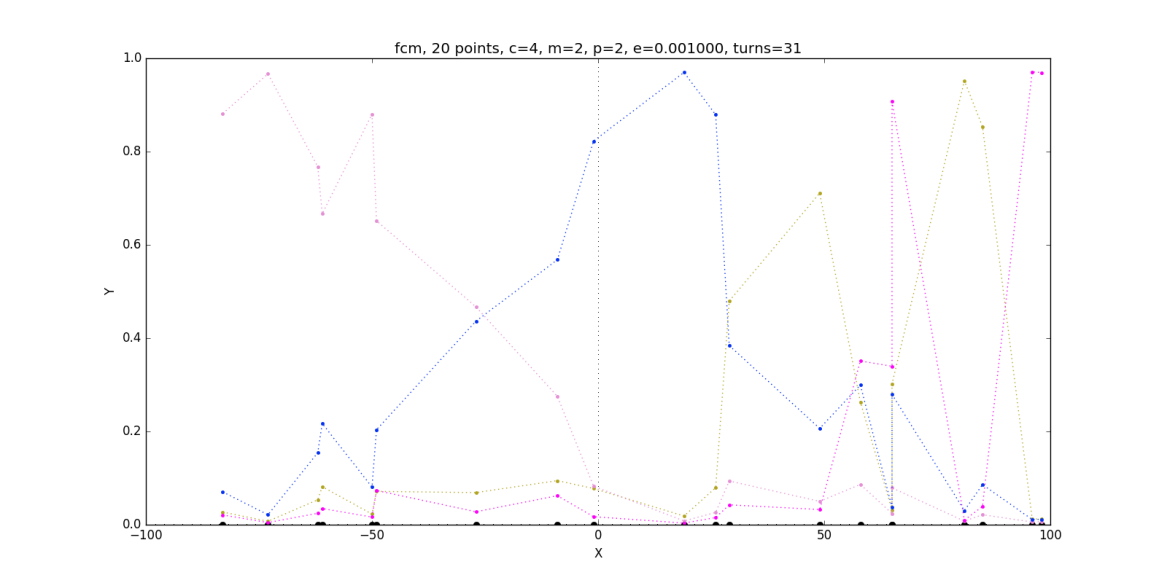
2可以简单的描述为针对不同的属性给予不同的权重值，决定其属于那个簇，，r,p为自定义的参数，根据实际情况选择，其中，p是用来控制各维的渐进权重，r控制对象间较大差值的渐进权重。当r=p时，即为闵可夫斯基距离，当p=r=1时为曼哈顿距离，当p=r=2时为欧式距离，当p=r并趋于无穷时即为切比雪夫距离(可以用极限理论证明).因此，这几种距离统称为闵氏距离，闵氏距离的不足在于：从横向(各维)看，它等同的看待了不同不同的分量，这种缺陷从切比雪夫距离中可以明显看出，忽略了不同维的差异。从纵向(单维)看，它忽略了不同维的各对象的分布差异，这种差异在统计学中可以用期望，方差，标准差等度量。

3使用matplotlib编写壳程序，维度取2（平面),p取2（欧式距离），结果图示如下。

200个点聚为4类，终止值0.001，运行了70轮。

  
  
  
20个点聚为4类效果。



上图的匹配度函数曲线效果（X轴投影）。  


陈师姐的做法是把100天的数据聚成八个类，每个类再按24小时聚类成4类；

分类或等级变量之间的相关性可以用Spearman秩相关系数，只要两个变量具有严格单调的函数关系，那么他们就是完全spearman相关的。

当样本的某些属性是分类变量时，均值可能无定义，可以使用K-众数方法。

聚类，离群点检测，预测，这三样分别对应三个成果。

第一步、数据清洗，删选，离群点检测

第二步、聚类，小波分解，最后用anfis来预测

我们还是用MATLAB加上origin，放弃使用Python编程，不要本末倒置。