**原版iterated AP cluster方法求初始中心**

即通过迭代AP cluster方法来实现聚类中心初始化

0 预设一个k值来确定求取的中心个数，这里k与初始化完成后的K-means/K-medoids聚类个数K不一定完全对应，K <= k

1 运行AP cluster

1.1 从全体样本中确定一个参考点（Reference point），这个点在整个算法运行过程中保持不动。求取的方式一般是用全体样本的grand mean。

【Grand mean: 将全体样本分成若干份（sub-sample），每份样本生成一个均值样本（即对每个feature求均值），这个均值为sub-sample mean；假如一共是10份sub-samples，那么就有10个sub-sample means；对这10个sub-sample means求均值，则可以得到“均值的均值”，即为grand mean。】

1.2 找出距离参考点最远的一个样本作为一个“假”类Si的中心ci；

1.3 对所有样本点，按照距离判断其距“假”的中心更近，还是距参考点更近，若距离“假”的中心更近，则将其归到Si中;

1.4 对中Si的样本求均值，获取新的中心c’i，若c’i与ci显著不同，则回到1.3继续执行；否则，将ci输出，并在全体样本中去除Si，回到1.2继续执行；当触发任一终止条件时算法终止，进行后处理；

【终止条件：（1）所有样本已全部聚完；（2）前t个“假”类的贡献率总和超过一定阈值，如50%；（3）第t个“假”类的贡献率低于一定阈值，比如1/N；（4）k值达到预设的大小】

（关于贡献率的计算方式我还没查到）

1.5 后处理：若得到的某个“假”类中包含的样本个数少于某个阈值，则剔除这个“假”类中心（这一步不一定要，可交由人工判别）；输出所有“假”类的样本个数、贡献率及其中心ci。

**调整iterated AP cluster方法以适应DTW距离和K-medoids**

与原版方法不同的地方会特别标记

0 预设一个k值来确定求取的中心个数，这里k与初始化完成后的K-medoids聚类个数K不一定完全对应，K <= k；预设Si更新中心的尝试次数；

1 运行AP cluster

1.1 从全体样本中确定一个参考点（Reference point），这个点在整个算法运行过程中保持不动。求取的方式采用全体样本的grand median。

【Grand median: 将全体样本分成若干份（sub-sample），每份样本选出一个“中位”样本，这个中位样本与sub-sample中的其他成员平均DTW距离最近；假如一共是10份sub-samples，那么就有10个“中位”样本；再求这10个sub-sample“中位”样本的“中位”样本，即为grand median】

1.2 找出距离参考点最远（采用DTW距离）的一个样本作为一个“假”类Si的中心ci；

1.3 对所有样本点，按照DTW距离判断其距“假”的中心更近，还是距参考点更近，若距离“假”的中心更近，则将其归到Si中;

1.4 采用K-medoids更新中心的策略来获取新的Si中心c’i，若c’i与ci显著不同（或尚未达到Si更新中心的尝试次数），则回到1.3继续执行；否则，将ci输出，并在全体样本中去除Si，回到1.2继续执行；当触发任一终止条件时算法终止，进行后处理；

【终止条件：（1）所有样本已全部聚完；（2）前t个“假”类的贡献率总和超过一定阈值，如50%；（3）第t个“假”类的贡献率低于一定阈值，比如1/N；（4）k值达到预设的大小】

1.5输出所有“假”类的样本个数、贡献率及其中心ci