多元正态：

### 因子分析(FA)

，，，，，为对角阵

，。。。。。

当时，无法进行FA，和不唯一

对变量的拉伸不影响FA的结果，即不影响（左乘对角阵）

PCA方法进行FA：当最后的个特征值很小，则可忽略其在整体方差中的贡献，有，，误差的元素平方和小于等于（未加上特征值平方和），该方法一般性因子所能解释的方差比例，因为（额外假设列正交，很小）

MLE方法进行FA：认为一般性因子和特殊因子都是正态分布，即，，且二者独立，还要满足。计算时通过联系到最大似然式子中，当进行标准化时，有，则相应地，。（额外假设：多元正态）

旋转可以提高因子的可解释性。Varimax旋转标准：拉伸，选择旋转，使得。估计因子得分：认为和已知，计算方法有WLS(加权最小平方)、OLS(一般最小平方)和regression(回归)

WLS：最小，可得（Bartlett分数）（、、为根据样本估计值）（标准化不影响得分估计）（该估计是无偏的）

OLS：，可得，其中（PC方法下）

Regression：认为，则，故而，故而（如果采用PC方法得到的和，则，与PC得分有关）

**在MLE方法下得到的和无论是否标准化都不影响得分**

，，WLS方法得到的估计的方差为，R方法得到的估计的方差为更小

发生旋转时，估计值仍满足，

### 典型相关分析(CCA)

找到一组和，使得维度数为和的两个随机变量的最大，。（）

第i对典型相关对：，且满足，且，，对于。

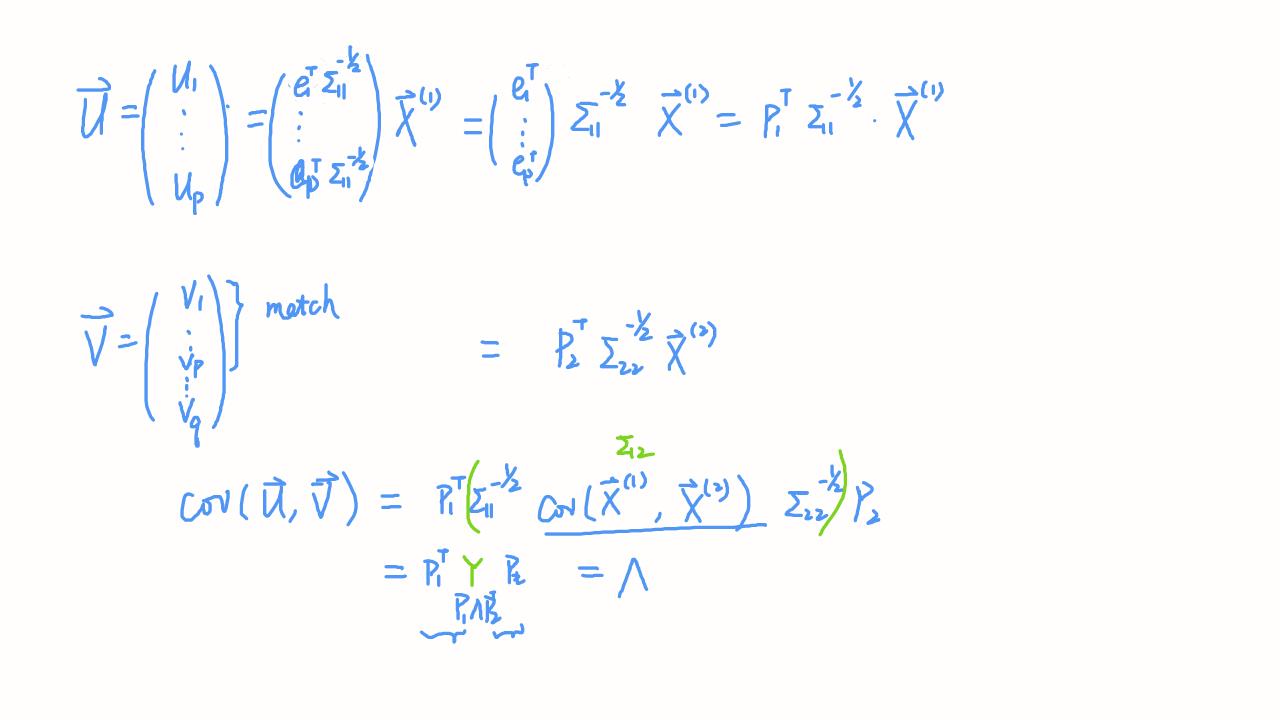
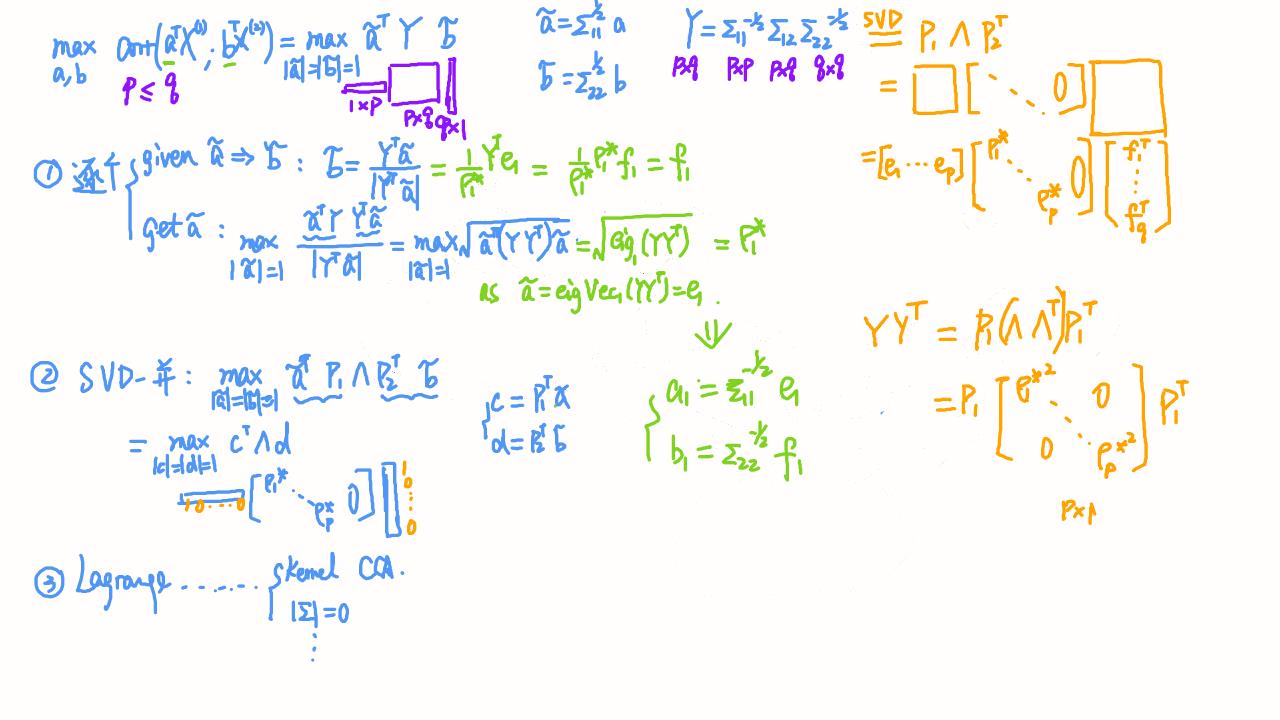
，，为的单位特征向量，为的单位特征向量，为第大的特征值，，。（右上为）

新加性质：。（注意相互不垂直）因为为对角元素为的对角阵。，。（此处二者维数都为）（，）

，，，。

，为被解释的方差比例，也为被解释的方差比例。

数据标准化（旋转平移）后，与和都不变。



### 判别分析(DA)

LDA：取两个总体的方差为，二者分离度为，让分离度的平方最大，即，该最大值为，而。（Fisher）

ECM：假设两个分类为、，两个分类出现的前验概率为、，错判代价为、。令最小，从而得到与的判定分界点：，反之。（与是两个总体的已知分布）

TPM：忽略错判代价的ECM，相当于。

Bayes：当，则认为其应被分为，，其结果相当于的ECM。

对两个多元正态分布的总体的分类（LDA）：①认为两个总体方差一致，则分界。（对于样本数据，取，，）

QDA：如果认为两个多元正态分布方差不相等，则（其中）。

对于两个等方差正态分布的最小TPM为（其中）

APER（显著错误比例）：尽量减小如此分类（确定界限后）后的错误比例APER，易出现过拟合（LOOCV: 对每个样本点x，用排除x的全 部数据作训练集，预测x的分类 ；n-fold CV: 将样本均分n份。对每份，用其他样本作训练集，该份作验证集。）

出现过拟合，可以采用Cross Validation：一部分数据用于训练，剩下数据用于检验

对全部数据进行整体标准化（线性变换），不影响分类结果。

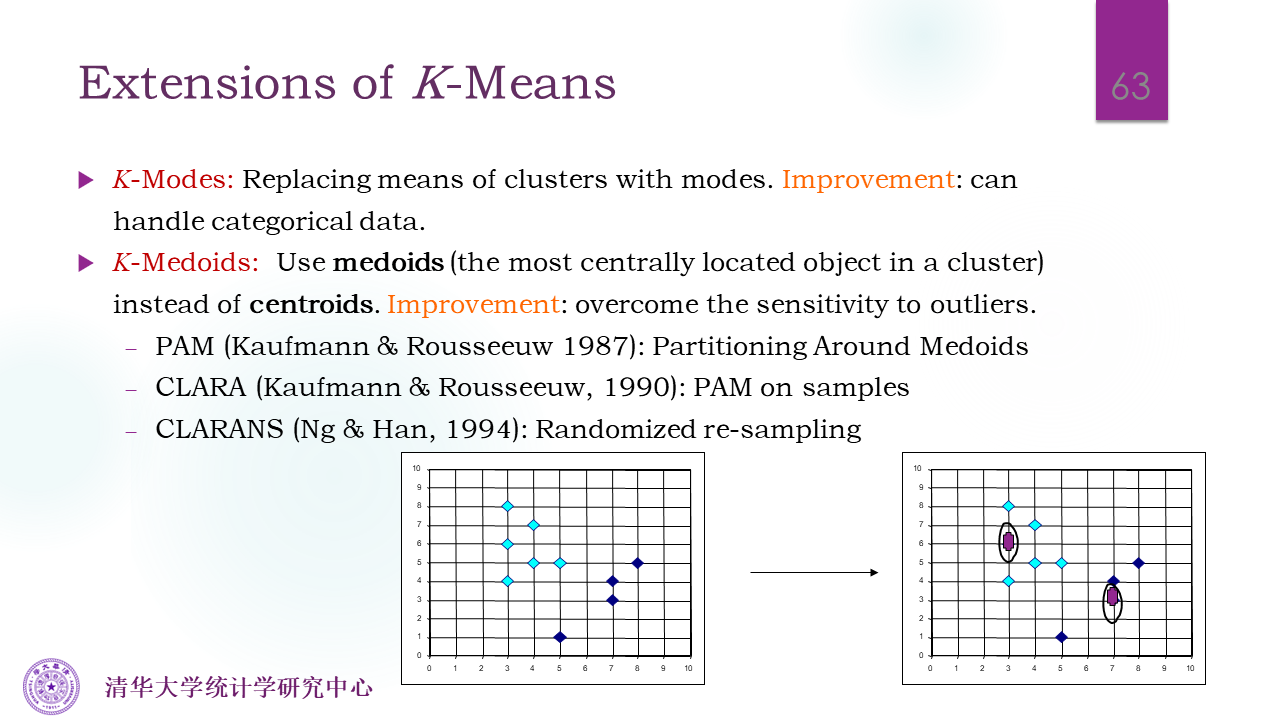
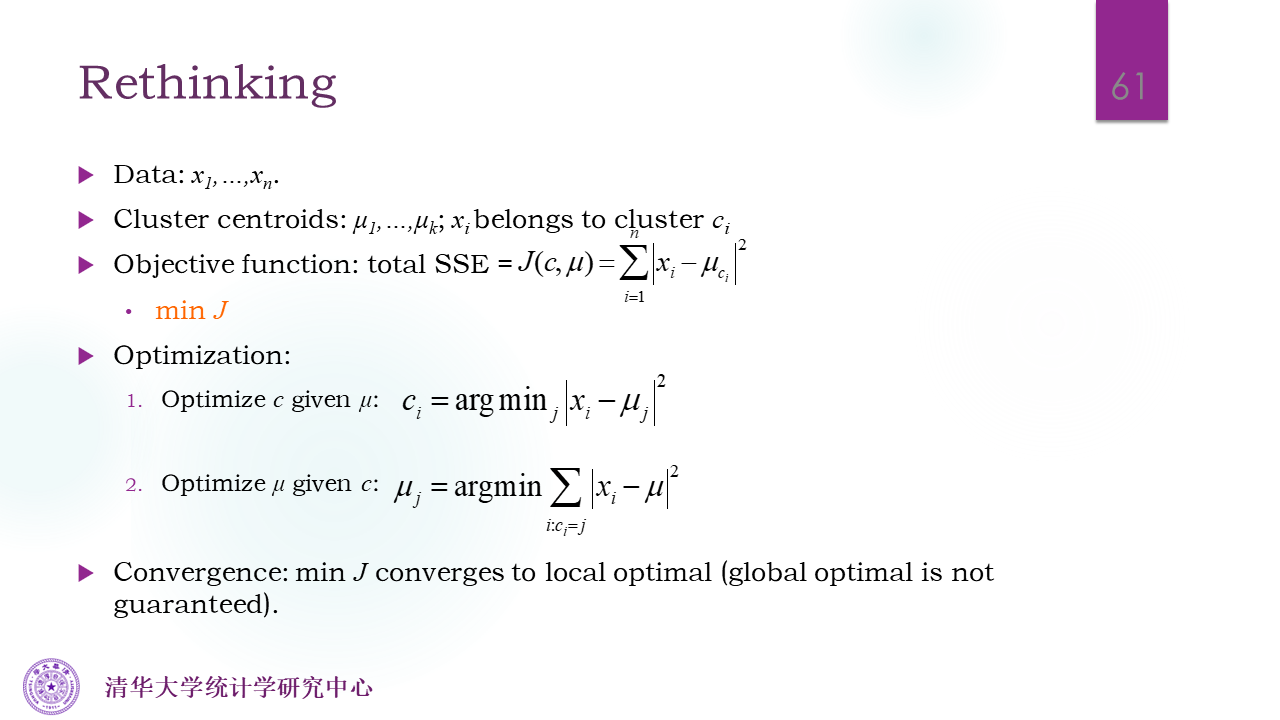
### 聚类分析(MSA)

**Hierarchical Clustering**（层次聚类法）：

Agglomerative 聚类：需要定义点与点和聚类与聚类的距离（single linkage、complete linkage、average linkage、Ward's linkage、质心）（Ward' linkage：，即从最开始合并时，尽量让分出的块的方差最小）（不需要提前确定分类数目，容易展现层次关系，但是计算复杂度高）

**Partitioning Clustering**：让每个数据被分到一个没有交叠的子集中

K-means聚类法：先给出初始K个质心，此后根据质心进行分类（离同一质心最近的分为一类），然后再重新确定各自质心，重复操作直到稳定。（需要实现确定分类数K，点点距离定义，初始质心位置）（确定K数的方法：①Elbow：让SSE不断大幅度降低，到达缓慢变化时的临界点取为K；②Empirical经验主义：；③Cross Validation：一部分分类一部分检验）（效率高，但对异常值敏感，且对奇怪形状分类有问题）



（可以把换为K-Modes或者K-Medoids，找到其中最接近中心的点，而非质心，这样就避免了对异常值的敏感）

**Model-based 聚类法：**(soft)

Mixture Modelling：认为某种观测值可以来自很多中类别分布的交叠。每种类别有自己的概率分布和权重。一般认为各个分类都为多元正态分布。，。目的就是估计概率和。

假设已知分类Y，那么就是个分类问题：，不断更新，使得，，。

实际上未知，所以。计算时在推测和推测中切换。此时