1. LASSO和岭回归对于线性回归中不可逆不能用最小二乘求系数或使用时造成系数过大的情况下（该情况因变量多重共线性造成的），加入LASSO（绝对值约束）或岭回归（二次约束）就可以求系数。LASSO直接选择变量，不需要重新组合变量，顶点处不可导，不能用求导求极值，用LAR来。岭回归只能消多重共线性不能做变量选择，可以用求导求极值。
2. 降为手段（太传统的方法）：主成分分析因子分析（需要重新组合变量），都只能处理线性问题。主成分分析本质就是坐标轴变换，选择方差大的，变换后选择的主成分可能是某些变量X的线性组合，故可解释能力不强，业务不太好说。因子分析是将每个X变为公共因子F的线性组合，如将数学，物理化学学习能力变为公共因子理科学习能力，业务上可解释可能比主成分分析强一点。
3. （KNN,K最近邻法（离最近的K个点属于哪一类，判为最多的那一类）【距离判别法】）

【线性判别法】

【贝叶斯分类器】

1. *先验概率：之前已经知道的概率。*
2. *朴素贝叶斯P(A/B)=P(B/A)\*P(A)/P(B) 其中P(B/A)=P(B1/A)\*P(B2/A)……..(文本分类：垃圾邮件过滤，新闻类型分类，用户流失)，特征是独立的*
3. *贝叶斯信念网络（BBN）*
4. 聚类：KMEAMS，K中心法对于球形图簇比较有效，非球形不是很有效，聚类中有基于密度的方法，DBSCAN，适用于非球形，

DBSCAN：R领域为给定半径为R的区域，核心点：如果一个点的R领域至少包含最少数目M个点，则称该点为核心点。直接密度可达：如果点P在核心点Q的R领域内，则称P是从Q出发可以直接密度可达。

思想：

1. 指定合适的R和M
2. 计算所有样本点，如果点P的R领域里有超过M个点，则创建一个以P为核心的新簇
3. 反复寻找这些核心点可以直接密度可达的点，将其加入到相应的簇，对于核心点发生“密度相连”状况的簇，给予合并
4. 当没有新的点可以被添加到任何簇时，算法结束
5. 大数据情况下聚类：
6. 抽样
7. 精度降低
8. 分布式

方法一：从大样本抽一部分，再用K中心法进行聚类，得到K个中心，然后把剩下来的样本离这K个中心距离，把样本分进去。

1. 基于网格的聚类方法：CLIQUE

用网格中点的稠密度来看网格是否是稠密的网格，再合并周边稠密的网格，形成簇。