1.决策树

2.朴素贝叶斯

3.最小平方回归

4.逻辑回归

5.支持向量机

6.集成学习

7.聚类算法

8.主成分分析

9.奇异值分解

10.独立成分分析

-----归纳偏执bias和过拟合问题----

机器学习的过程类似于归纳，利用以前的一些特征来归纳出一定的规则，遇到新的物体的时候在该物体上使用该规则，一个好的算法需要一定的偏执，用于防止过拟合现象。

我们学习到的所有的算法主要是适用于新的事物，当没有偏执的时候我们总能拟合出过所有样本点的曲线，但是这并不适合与新的样本点

------机器学习中模型的大的分类----

主要分为概率形式和非概率形式

概率形式-----类似于logisit回归使用softmax/sogmoid函数求得每个样本X属于Y的概率

概率形式的模型主要分为 生成式和判别式

一个是直接后验概率进行求解，另外一类是转换成联合分布来进行求解

非概率形式：通过y=fx 直接将X映射到y上面

二、聚类

-------kmeans----------------------------

Kmeans聚类主要依据距离的衡量来进行聚类----同一类样本之间的距离很小

算法流程：

　　　首先选择组个数，每组的组中心

　　　利用之前迭代的组中心将每一个样本进行分类（求解每一个样本到组中心的距离，将其划入距离最小的一类）

　　　然后利用每一个组的样本更新组中心

　　　不断重复上述两个步骤，直到容忍度

Kmeans也算EM算法的一个变种

-----em算法 处理优化问题----------------------------

估计+最大化

求解似然函数的参数的一种新方法

EM是一种解决存在隐含变量（分类变量）优化问题的有效方法

问题：给数据集进行分类

思想：EM算法求取参数主要是不断地极大化对数最大似然函数（联合分布的）的下界，来求解对数似然函数，从而求得参数 （下界主要使用Jensen不等式进行放缩 log(x)是凹函数）

算法：假设出样本集合和类别变量z之间的联合分布的参数

利用后验概率求得每个样本属于每个类别的概率（E步）

然后对似然函数的下界进行最大化 （M步）

最后输出联合分布的参数

最后的似然函数（此处的似然函数为原来对数似然函数的下界）在t时刻优化得到的最大值即为t+1时刻的似然函数的下界。

似然函数单调增加，最后得到原来似然函数的最大值，即对应找到联合分布的参数

------高斯混合模型聚类法--------------

高斯混合模型一般用于聚类以及密度估计

基于后验概率进行分类--主要是假设数据集是服从高斯分布，首先假设出总的组数K，即所有的数据集选自k个高斯分布，利用em算法来求得每个高斯分布的参数均值、方差以及每个高斯分布出现的概率

和kmeans是一样的不断的去更新每个组的质心，直到最后达到收敛

一般求得每个类别的高斯分布的参数 只要使用最大似然函数（或者对数似然函数）（混合模型的）进行求最大值即可求得，处于计算难度的考虑使用以下算法进行不断更新迭代出合适的参数

算法流程：

　　　首先输入总的类别个数K，以及假设每个高斯分布的参数均值和方差

　　　利用上一轮迭代求得的均值和方差--求取每个样本属于每个类别的概率（混合模型的后验概率）

　　　对于每个类别生成一列新的样本(可以看做是这样)-----从EM算法对原来似然函数的下界进行求导得到的思路-----求导得出均值的估计为以下样本序列的均值，方差为该序列的方法

　　　R(i,k)表示每个样本属于第k个类的概率

　　　利用每个类别新的样本点进行求解每个高斯分布的均值和方差

　　　不断地迭代上述两个步骤最后找到每个类别的均值和方差，从而最后求得每个样本的后验概率（基于混合模型）---以此来进行分类

　　　（与kmeans是类似的不断去更新组中心，不过gmm方法主要使用高斯分布来获取每个类别的组中心以及每个类别的组方差）

-------层次聚类法------------

层次聚类区别于Kmeans的一点是，层次聚类法不提前设置组个数以及组中心，层次顾名思义是逐层聚类，主要分为两种方法：分裂法和凝聚法

分裂法：

　　　开始将所有的样本看成一个大类，然后按照距离最大 法进行分类，将所有样本距离最大的两个样本a,b分类到两个大类C1,C2中，

　　　然后对于剩余的样本与之前的两个样本进行求解距离，将其进行分类(如d与a的距离更近将其分到C1类中)

　　　将所有的样本分类好了之后继续对上述两个大类分别使用之前的方法进行分裂，直到达到相应的组数。

凝聚法：和分裂法正好相反是开始时将所有的样本分别看成一个类，然后挑选距最近的样本聚合，直到达到想要的分类组数

------谱聚类----------

谱：方阵作为线性算子，该方阵的所有特征值统称为谱，方阵的谱半径为最大的特征值。矩阵A的谱半径是矩阵的最大特征值。

　　　谱聚类的方法主要哦是将原来的聚类问题转换成图聚类问题，首先是将原来的问题转化成一个图，把每一个样本看作图中的顶点，然后利用相似度分析只将其中相似度最高的K个顶点的边进行相连，这样就建立一个图。谱聚类的思想主要是希望同类之间的相似度高，异类之间的相似度低。将问题转换为一个图之后，只需要进行图分割。由于图分割的标准有很多，因此谱聚类将图划分准则优化问题转换成求解相似矩阵或者Laplace矩阵特征问题，可将此类方法统称为谱聚类方法，也可以认为谱聚类方法是对图划分准则的逼近。

谱聚类是将样本从高维空间中映射到低维空间，然后在低维空间中进行kmeans聚类，降维的主要思想主要类似于PCA的方法通过选择最大的K个特征值进行降维

但是这个地方的特征值为数据集的laplace矩阵的特征值

构造的时候将数据集中的每个样本点看作无向图中的定点，图中的任意点之间可以相连也可以不相连，任意两个顶点之间的权重为两个样本之间的相似度

Laplace矩阵的构造：

　　　首先获得样本的相似矩阵S（距离矩阵），然后获得权重相邻矩阵W（该矩阵依据S矩阵进行求解），然后构造D矩阵，利用W矩阵求得每个顶点的（和vi顶点相连的那些边的权重之和），然后构造Laplace矩阵L=D-W，其中D,W分别为对称矩阵，而且D为对角线矩阵，对角线的元素存放。

优化规则：RatioCut切图、Ncut切图

谱聚类切图：主要利用谱聚类的主要思想

https://www.cnblogs.com/pinard/p/6221564.html

三、奇异值分解

　　奇异值分解主要是用来求解非方阵的特征值和特征向量问题，在推荐系统中，奇异值分解起了很重要的作用，在面对用户-项目 矩阵的分解问题上提供了很大的便利，奇异值分解，将该矩阵分解为两个矩阵，其中一个矩阵为关于用户的偏好矩阵，另外一个为item的特诊矩阵，然后利用这两个矩阵去预测原来的用户-项目矩阵，从而达到推荐的效果

--------白化预处理--------

白化是指：对于一个零均值的随机变量来说，其方差为单位矩阵。

白化预处理过程---将混合数据进行白化预处理之后，得到的样本和源信号一样都是独立的，减少了ICA的工作量。进行白化之后拿到的观测数据的各个分量之间是独立的且方差一致。利用新得到的数据再进行ICA分析

https://blog.csdn.net/whiteinblue/article/details/36171233

--------Laplace分布-------

--------峰度--------------

四、独立成分分析

　　　注意：在ICA源信号中的多个源信号的分布至多一个是高斯分布，因为高斯分布在进行线性变换之后还是高斯分布，这样混合信号总是独立的，这样求解结果就不是唯一的了。而且一般在进行ICA分析之前总是要进行白化预处理。

　　　ICA常用于信号去燥的过程中，从多个维度的数据中获得有用的数据，在这个过程中并没有实现纬度的下降。

　　　常见的信号处理问题：在房间的不同位置放置n个收音器，获得n个声源的信号这样组成一个n维的信号样本，在T个时段分别进行检测可以得到T个这样的样本。在现实世界中我们只能接受到加固噪音之后的声源，源信号和收音器材的本身的干扰信息是不知道的，因此我们需要从混合声源中获得源信号。但是在只有混合声源的情况下求取源信号和收音器材的干扰信息。因此我们需要先根据混合数据的峰度来确定出源数据的数据分布（若峰度大于0则是超高斯---laplace分布，小于0均匀分布），一般不知道的情况下，可以假设源信号的服从sigmoid函数的。

　　　ICA在处理信号去噪时，我们的数据的分布是非高斯分布（通过数据集的峰度进行检验）的而且源信号之间是相互独立的，然后利用极大似然估计来估计收音器材的干扰权重。

　　　与IA,PCA的区别：

　　　IA----因子分析法，是一种降维是手段，但是和PCA不一样的是，因子分析法挖掘出的是能够表示众多原始数据的低纬度的数据。低纬度的数据的变量是高纬度数据的变量的成因。

　　　PCA ----去无关特征的过程

　　　ICA ----解混合信息的过程 并未降维

五、支持向量机

　　　以下均以二分类讨论

　　　1.从logisit回归到支持向量机-------在logisit回归中对于二维分类情况来说，有两个概率，分别是为正例的概率以及反例的概率，当正例的概率大于0.5时可以推得超平面，反之为为反例。因此我们在进行线性分类的时候主要是找到一个超平面，在该超平面的一侧均为正例，另外一侧为反例。即寻找的超平面为。

　　　2.函数距离 求数据平面上的任意一点到超平面的距离，过点x存在一条和超平面相平行的平面，则为函数距离，而在超平面的一侧的数据的分类变量y的符号和h(x)是一致的因此。

　　　3.几何距离：因为w和b是同时变化的，若超平面为时，函数距离变大，但是此时超平面以及数据点均为变化，因此使用几何距离.

　　　4.超平面不固定 给定一个w和b产生一个超平面，因此我们的优化目标是在所有几何距离最小的点中找到距离最大的点。

　　　5.函数距离的不固定，由于对于支持向量来说存在一条直线，当w和b变成原来的2倍的时候，函数距离也发生变化成为原来的二倍，但是我们的超平面和平面中的点均未发生变化因此对于是支持向量这些点来说让，则对于所有分类正确的点来说.

通过上述6点则可以找到我的优化目标：

　　　6.拉格朗日求解：在进行拉格朗日求解原来的优化问题，通过分析，因此。之后需要注意的是该问题的对偶形式的证明。从是原约束问题的下界开始，则为原约束问题的下确界。因此

　　　7.求解使用smo算法

　　　8.硬间隔和软间隔---硬间隔要求所有的样本必须处于支持超平面的两侧，不能在间隔之内存在，而软间隔则放宽了限制，可以允许样本点穿过支持超平面，甚至是超平面。

9.在软间隔情况下函数距离发生变化。 其中是松弛变量。但是不能无限大，这样的话对于任意的超平面都满足上述要求了，则存在一个惩罚因子C，C越大说明离群点对目标函数的求解影响越大。则目标函数变为：

10.非线性分类与核函数：在 低纬空间不线性可分的样本映射到高纬空间中可以会线性可分，因此出现了核函数的存在，即存在此映射。其余和线性的情况是一致的。核函数即.