***Week 15***

***优化方法，数据中心化，奇异值分解，软阈值收缩，基于PCA的人脸识别算法，*通过正则化自表达的无监督特征选择**

**机器学习常用优化方法**

**无约束：  
1. 梯度下降法**

思想：梯度下降是常用的一阶优化方法，用来求解无约束优化问题。用当前位置负梯度方向作为搜索方向，该方向为当前位置的最快下降方向，也称最速下降法。最速下降法越接近目标值，步长越小，迭代越慢。当目标函数为平滑凸函数时，梯度下降法的解为全局最优解。其它情况不能保证全局收敛，因此调参很重要。当目标函数二阶连续可微时，可使用牛顿法。其每轮迭代次数远小于梯度下降，但每轮迭代都需要对海森矩阵求逆，计算代价很高。于是以较低的代价寻找海森矩阵的近似逆矩阵，由此引出拟牛顿法。

缺点：

（1）靠近极小值时收敛速度减慢；

（2）直线搜索时可能会产生问题；

（3）可能会“之”字形下降。

算法：给出m个样本，设估计函数为，令，则可以表示为。 损失函数为。

先求偏导：；

然后更新：，（为学习率）。

基于基本的梯度下降，衍生出两种方法：随机梯度和批量梯度。

批量梯度下降：最小化所有样本的损失函数，最终求得全局最优解，但面对大规模样本时效率低下。

随机梯度下降：最小化每个样本的损失函数，每次迭代不一定朝着全局最优方向，但最终结果在全局最优解附近，适用于大规模样本的情况。

1. **牛顿法和拟牛顿法**

牛顿法：，又称切线法。

1. **共轭梯度法**
2. **启发式算法**
3. **近端梯度法**

如果目标函数在某些点不平滑，则这些点梯度无法求解，传统梯度下降方法无法使用，于是引入近端梯度下降方法。思想是用近端算子作为近似梯度来进行梯度下降。

1. **加速近端梯度法**

通过加入上一次迭代方向的一部分，使本次迭代方向与上次方向不至于偏离太远，达到加速的效果。

**有约束：**

1. **拉格朗日法**
2. **二次规划**
3. **半正定规划**

**数据中心化：得到的每一个样本减去全体样本的均值。**

**数据标准化：**将数据按比例缩放，使之落入一个小的特定区间。在某些比较和评价的指标处理中经常会用到，去除数据的单位限制，将其转化为无量纲的纯数值，便于不同单位或量级的指标能够进行比较和加权，**使不同特征对结果的贡献相同。当原始数据在各个特征上尺度不一致，且模型处理结果具有伸缩不变性时，需要对数据进行标准化处理。**

**标准化的具体方法：**

min-max,zscore,归一化

**基于PCA的人脸识别**

*GBKII*

介绍：基于Grey距离的KNN迭代填充算法，在计算GRG（i,j）时要求实例i和j有相同类标签的假设下，用GREY距离替代欧式距离，可以减少时间复杂度。另外，可以克服EM算法缓慢的收敛速率。

相关工作：MI算法重复独立填充M次，EM算法基于参数模型重复地交替。GBKII在第一轮填充时用观测到的属性值的均值填充缺失值，从而能最的地利用全局信息。从第二轮开始，迭代填充是基于上一轮的填充结果。GBKII是类似于EM的迭代填充，但是是一个非参算法。

GBKII：

很多算法用无缺失值的实例作为参考实例，然而数据库中无缺失值的实例很少。有的实例虽然有缺失值，但仍包含了大量信息。因此利用所有实例来填充缺失值很有必要。

但是我们无法利用全局信息，因为有缺失值。所以首先用属性均值来填充缺失值。在机器学习和统计中用均值填充是流行且合理的。有人认为当且仅当数据集是从正态分布的总体中选取时，均值填充才会有效。然而现实应用中，无法提前得知数据集的真实分布。

收敛性：最快。

**通过正则化自表达的无监督特征选择**

**摘要：**

通过移除不相关的冗余特征，特征选择意在寻找一个关于原始特征的有良好泛华能力的紧凑的自表达。被在子空间聚类中的低秩自表达的成功所启发，我们为无监督特征选择提出了一个正则化的自表达模型，其中每个特征可由相关特征的组合线性表出。通过使用范数来构造自表达系数矩阵和自表达残差矩阵，RSR可有效选择表达有代表性的特征且能保证对于异常值的鲁棒性。如果一个特征很重要，它会参与到大部分其它特征的表达中，反之亦然。

正如稀疏性导致了稀疏表达，自相关性导致了自表达。

在这篇文章中，特征矩阵被自身表达来寻找有代表性的特征组成。自表达残差通过范数损失函数最小化来减少异常值的影响。

**问题描述：**

无监督特征选择的目标是从无标签信息的给定数据集中选出期望的特征子集。

行向量为样本，列向量为特征。一个健壮且优秀的特征选择算法应该能去除异常值的影响并指示出冗余特征。

首先计算样本相似度或流形结构，然后构造一个响应矩阵，特征选择问题可转化为一个多重输出回归问题：

（1）

其中是特征权值矩阵，是施加在上的正则项。方程（1）中，响应矩阵在优化阶段之前已知，是变量。包含了样本相似度的信息，且在不同方法中的计算方式不同。

正则化自表达：

1. 中模型同时考虑了样本相似度以及选择特征。响应矩阵的选择很困难，由于特征的自表达属性，我们提出了正则化的自表达模型RSR。RSR简单地将作为响应矩阵，可以被自表达原则很好地解释。对中每个特征，用其它特征的线性组合来表示：

（2）

于是对于所有特征：

（3）

应该反映不同特征的重要性，同时使表达残差最小。F范数可用来度量残差，但对异常值敏感。考虑到一个异常值样本是的一行，一个表达残差是的一行，我们用范数来构造E；也就是对E施加行稀疏来实现对异常值的鲁棒性。同时，如果使，可以得到平凡解。为了避免平凡解，引入正则项,于是有了如下的最小化问题：（4）

可以作为特征权值，因为它反映了第个特征在表达中的重要性。我们让，方程（4）就成了

（5）

称以上模型为对于无监督特征选择的正则化自表达模型。

**优化和算法：**