**简答题**

1. 简述PCA的原理、学习模型和算法步骤

**原理**：PCA主要是想寻找一些方差最大的方向，将数据沿着这些方向进行正交投影，通过这样的投影，可以尽可能使得投影后的样本点较分散，即投影后的方差大。投影后样本的方差可以表示为，由于是正交投影，有约束条件。使用拉格朗日乘子法求解最大化投影后样本方差的问题，描述为，其中是原样本的协方差阵。令其偏导等于0，可以得到。即投影向量为方差矩阵的一组表征正交特征向量。

**学习模型**：设样本点是零均值化的，是投影后的协方差。根据PCA的原理，学习模型可以表示为：

**算法步骤**：

1. 将数据零均值化：，
2. 计算数据的协方差阵：
3. 对矩阵C进行特征值分解，选取最大的m个特征值所对应的标准正交特征向量，则投影阵
4. 对数据进行投影：
5. 简述LDA的原理和学习模型

**原理**：LDA是想寻找一组投影方向，使得样本经过投影后，同类样本尽可能地靠近，不同类的样本尽可能地相互远离。以两类分类问题为例，对于同类样本投影点尽可能地接近，可以使用投影后样本协方差阵比较小，对于不同类样本尽可能地互相远离，可以让类中心点地距离尽可能地大，即大。由此可以变成最大化。记类内散度矩阵为

类间散度矩阵为。问题变成最大化，约束条件为。对问题进一步表征，。对此问题应用拉格朗日乘数法进行求解，，因此向量w为矩阵的的特征向量。

**学习模型**：对于c类样本数据，记类内散度矩阵为，式中

。

类间散度矩阵。

学习模型可以表示为以下两种：

1. 作为一类非线性降维方法，简述流形学习的基本思想

流形学习的基本思想为：高维空间中相似的数据点，他们映射到低维空间的距离也是相似的。

LLE算法保持了高低维空间中，局部区域中心点的线性重构关系。

Isomap算法保持了高低维空间中，任意两个点对之间的测地距离。

LE算法保持了高低维空间中，局部区域点对之间的亲和度。

LSTA算法对每一个数据，在局部引入一个线性变换，将其近邻点映射到低维坐标系中的对应近邻点。

LSE算法对每一个数据，在局部引入一个非线性变换，将其近邻点映射到低维坐标系中的对应近邻点。

1. 根据特征选择和分类器的结合程度，简述特征提取的主要方法，提出各类方法的特点

主要有三种方法：过滤式、包裹式以及嵌入式特征选择方法。这三种特征选择方法与分类器的结合程度不断加深。

**过滤式特征选择**

主要方法：特征选择过程与分类器学习过程没有任何关系。首先定义一个评价函数，并用它来度量某个给定特征与类别标签之间的相关度；最后选取具有最大相关度的m个特征作为选择结果。

基于过滤式的特征选择有：单独特征选择法、顺序前进特征选择法、顺序后退特征选择法、增l减r特征选择法、Relief方法。前四种是一种遍历求解的方法，即根据给定的特征评价函数，按照一定的增减特征顺序，对所有样本的特征遍历求得局部最佳特征集合。Relief一方面是在数据集的采样上进行，另一方面设计了一个相关统计量来度量特征的重要程度。

过滤式特征选择特点：

1. 过滤式方法先对数据集进行特征选择，然后再训练学习器。特征选择过程与后续学习器无关；
2. 启发式特征选择方法，无法获得最优子集；
3. 与包裹式选择方法相比，计算量降低了很多。

**包裹式特征选择**

主要方法：这是一种以分类性能为准则的特征选择算法，特征的选择依赖于分类器的结果。先对数据集进行特征选择，然后再训练分类器；特征选择过程与分类单独进行，特征选择评价判据间接反应分类性能。

基于包裹式的方法分两类：

1. 直观方法：这种方法给定特征子集，训练分类器模型，计算分类器错误率为特征性能判据，进行特征选择
2. 递归方法：首先利用所有的特征进行分类器训练，然后考查各个特征在分类器中的贡献，逐步剔除贡献小的特征。如R-SVM，SVM-RFE，Adaboost

包裹式特征选择特点：

1. 特征选择过程与分类性能相结合，特征评价判据为分类器性能。对给定分类方法，选择最有利于提升分类性能的特征子集，最终分类性能比较好。
2. 特征选择过程中需要多次训练学习器，因此包裹式特征选择的计算开销比较大。

**嵌入式特征选择**

主要方法：在学习器训练过程中自动地进行特征选择，对分类器的训练模型加入一定的限制条件，让权重变得稀疏，稀疏权重对应位置的特征可以当作被剔除，达到特征选择的效果。

基于嵌入式的特征选择常采用L1、L2范数，即将分类器优化函数加上权重1范数或者2范数的约束。

特点：

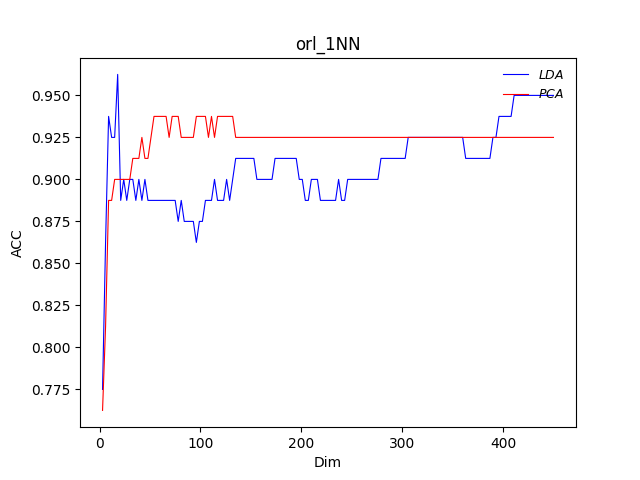
1. 将分类器学习与特征选择融为一体，分类器训练过程自动完成了特征选择。
2. 训练出的分类器权重比较稀疏。
3. 分类器防止了过拟合，提高了模型的泛化能力。
4. 简述最优特征提取的基本思想
   1. 穷举法：从给定的特征中，遍历挑选出最优的特征子集。如果有n个特征，需要挑选出d个特征，则有种特征组合方式，计算量十分巨大。
   2. 分支定界方法：将所有可能的特征选择组合以树的形式进行表示，树的每一层按照特征评价判据从左到右升序排列，当前层某一个结点判据值不会低于该节点的后继结点。进行搜索时，每次从右边开始搜索(右边的判据值大)，记录当前最大的判据值，不断回溯，回溯时如果某结点值小于最大值，就不用探索该节点的后继。由此搜索过程可以尽早达到最优解，而不必搜索整个树。

**编程题**

ORL数据，缩减维度到3~450，步长为3，对测试集分类性能ACC如下()：

|  |  |  |
| --- | --- | --- |
| 目标维度 | LDA | PCA |
| 3 | 0.775 | 0.7625 |
| 6 | 0.8625 | 0.8125 |
| 9 | 0.9375 | 0.8875 |
| 12 | 0.925 | 0.8875 |
| 15 | 0.925 | 0.9 |
| 18 | 0.9625 | 0.9 |
| 21 | 0.8875 | 0.9 |
| 24 | 0.9 | 0.9 |
| 27 | 0.8875 | 0.9 |
| 30 | 0.9 | 0.9 |
| 33 | 0.9 | 0.9125 |
| 36 | 0.8875 | 0.9125 |
| 39 | 0.9 | 0.9125 |
| 42 | 0.8875 | 0.925 |
| 45 | 0.9 | 0.9125 |
| 48 | 0.8875 | 0.9125 |
| 51 | 0.8875 | 0.925 |
| 54 | 0.8875 | 0.9375 |
| 57 | 0.8875 | 0.9375 |
| 60 | 0.8875 | 0.9375 |

上表仅展示了降至3~60维时预测准确度，下图为acc性能图



Vehicle数据，缩减维度2~11，步长为1，对测试集分类性能ACC如下：

|  |  |  |
| --- | --- | --- |
| 目标维度 | LDA | PCA |
| 2 | 0.652941 | 0.541176 |
| 3 | 0.664706 | 0.617647 |
| 4 | 0.688235 | 0.676471 |
| 5 | 0.682353 | 0.682353 |
| 6 | 0.688235 | 0.682353 |
| 7 | 0.670588 | 0.652941 |
| 8 | 0.694118 | 0.688235 |
| 9 | 0.735294 | 0.688235 |
| 10 | 0.729412 | 0.688235 |
| 11 | 0.694118 | 0.676471 |

