

# 第一部分

## 1

- PCA原理

寻找一组方差较大的方向，将原始数据（样本）在该方向进行投影。即，将数据在新坐标系下进行表示，保留少数在方差最大方向上的投影，达到数据变换、尽可能地保留原始数据信息和降维的目的。

- 学习模型

已知：给定 $n$ 个样本 $\{\mathbf{x}_i, i = 1, 2, \dots, n\}$ ，其中 $\mathbf{x}_i \in R^d$ ；

目标：选择 $m$ 个( $m < d$ )能最大化投影方差的方向进行降维，且 $\{\mathbf{w}_k, k = 1, 2, \dots, m\}$ 之间互相正交且为单位向量；

最终优化目标为：

$$\max \frac{1}{n} \sum_{i=1}^n (\mathbf{w}_1^T \mathbf{x}_i - \mathbf{w}_1^T \bar{\mathbf{x}})^2$$
$$s. t. \mathbf{w}_1^T \mathbf{w}_1 = 1$$

- 算法步骤

步骤一

计算数据均值

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$$

步骤二

计算数据的协方差矩阵

$$\mathbf{C} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

步骤三

对矩阵 $\mathbf{C}$ 进行特征值分解，并取最大的 $m$ 个特征值 ( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ ) 对应的特征向量  $\{\mathbf{w}_k, k = 1, 2, \dots, m\}$ ，组成投影矩阵 $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m] \in R^{d \times m}$

步骤四

将每一个数据进行投影： $\mathbf{y}_i = \mathbf{W}^T \mathbf{x}_i \in R^m, i = 1, 2, \dots, n$

---

## 2

- LAD原理

寻找一组投影方向，使样本在投影之后（即在新坐标系下）类内样本点尽可能靠近，类间样本点尽可能相互远离，提升样本表示的分类鉴别能力。其中，投影方向数小于原始数据的维度，因此投影样本即相当于将样本在子空间内进行表示，从而达到降维的目的。

- 学习模型

已知：样本集  $D = \{(\mathbf{x}_i, y_i), i = 1, 2, \dots, n\}$ ,  $y_i \in \{0, 1\}$ ; 其中, 令  $\mathbf{X}_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i$  分别表示第  $i \in \{0, 1\}$  类的示例集合、均值向量和协方差矩阵。

目标：同类样本投影点的协方差尽可能小，不同类中心点之间的距离尽可能大；

即

$$\begin{aligned} \max J &= \frac{\|\mathbf{w}^T \boldsymbol{\mu}_0 - \mathbf{w}^T \boldsymbol{\mu}_1\|^2}{\mathbf{w}^T \boldsymbol{\Sigma}_0 \mathbf{w} + \mathbf{w}^T \boldsymbol{\Sigma}_1 \mathbf{w}} \\ s.t. \quad \mathbf{w}^T \mathbf{w} &= 1 \end{aligned}$$

---

### 3

流形学习基本思想

高维空间相似的数据点，映射到低维空间距离也是相似的，因此可以寻找一个映射  $f: R^d \rightarrow R^m, m < d$ , 使得样本在高维空间中的关系仍然能够在低维空间得以保持。

---

### 4

根据特征选择与分类器的结合程度：

- 过滤式特征选择方法：“选择”与“学习”独立

- 主要方法

- 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；

- 方法特点

- 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；
    - 启发式特征选择方法，无法获得最优子集；
    - 与包裹式选择方法相比，计算量降低了很多。

- 包裹式特征选择方法：“选择”依赖“学习”

- 主要方法

- 直观方法：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。

- 每次分类器训练和错误率计算需要的计算量大，不适合大量尝试不同的特征组合。

- 替代方法(递归策略)：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。

- 方法特点

- 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。

- 通常采用交叉验证来评价选取的特征子集的好坏

- 嵌入式特征选择方法：“选择”与“学习”同时进行

- 主要方法

- 对于线性分类方法  $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$ , 在学习  $\mathbf{w}$  时，对  $\mathbf{w}$  进行限制，使其不仅能满足训练样本的误差要求，同时使得  $\mathbf{w}$  中非零元素尽可能少。

- 方法特点

- 不能直接设置最终选择特征的个数 $m$ ;
- 通过设置正则化系数 $\lambda$ 来隐式控制 $m$ ;
- $\lambda$ 值越大, 模型越关注稀疏性, 得到的非零系数个数越少; 反之, 非零稀疏个数越多;
- 可以设置一个选择特征个数的上限, 通过设置不同 $\lambda$ 值, 得到满足要求的特征。
- 是一种嵌入式特征选择方法: 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。

## 5

最优特征选择的基本思想就是从给定的特征集合中, 选择与任务相关的特征子集, 去除与任务不相关特征、冗余特征、或者关联性较小的特征, 降低学习任务的难度, 提高分类器性能。

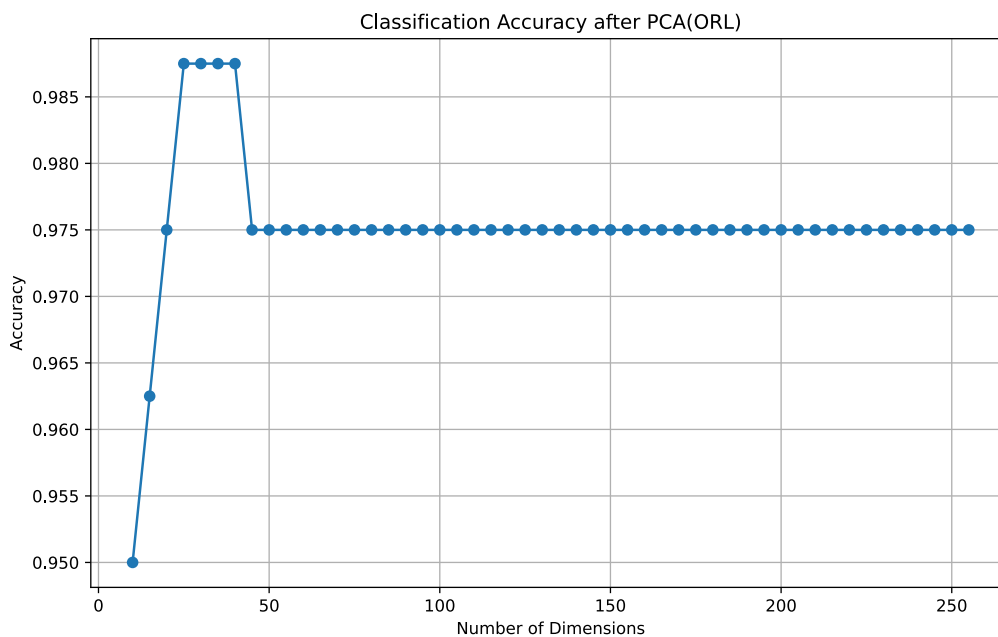
总体技术方法分为两个步骤, 即子集搜索和子集评价。构建特征的评价指标 (根据分类错误率、距离度量等), 利用穷举、分支界定等方法搜索最优特征, 得到最优的特征子集。

## 第二部分

(a)

PCA+KNN

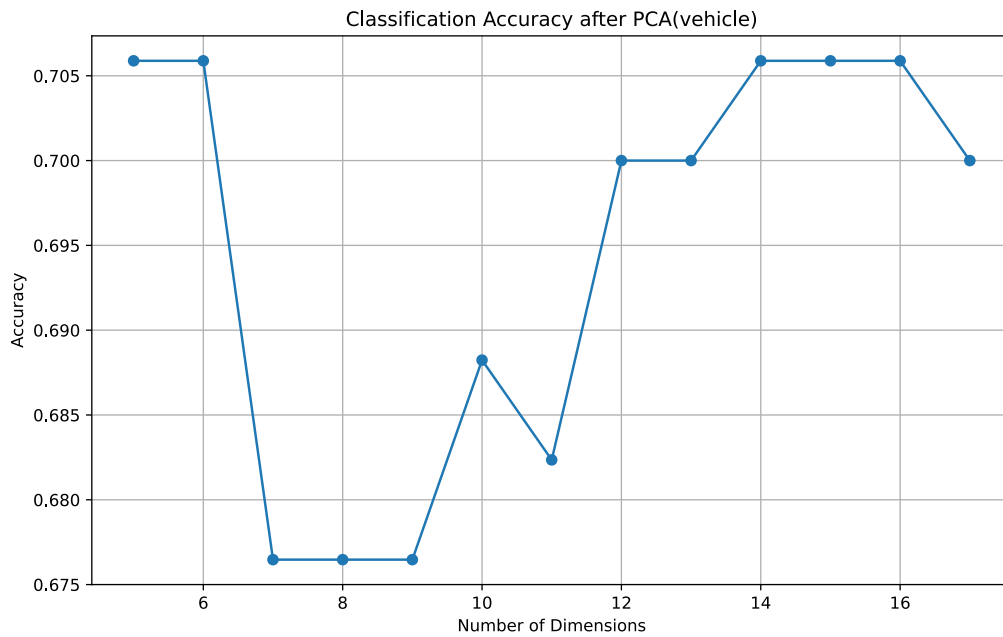
其中AT&T数据集结果如下图所示



当PCA降维的维数为[25, 30, 35, 40]时, 准确率达到最大值98.75%;

准确率随着降维维数不断增加时, 准确率先提高后下降, 最终收敛至97.50%;

Vehicle数据集结果如下图所示



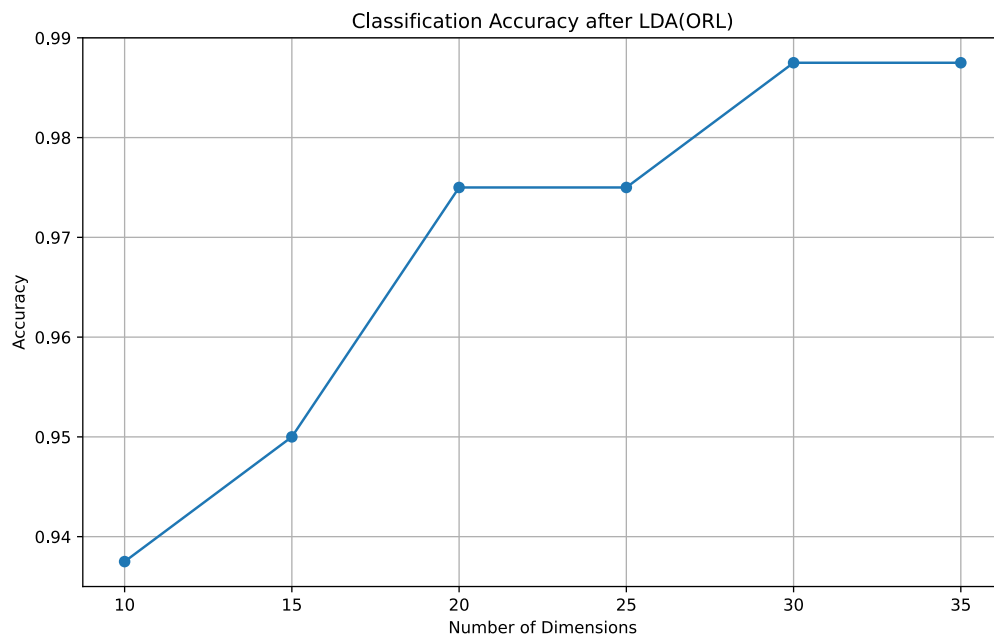
其中降维维数设置为[5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]

其中降维维数为5, 6和14, 15, 16时, 准确率最高, 为70.59%;

(b)

LDA+KNN

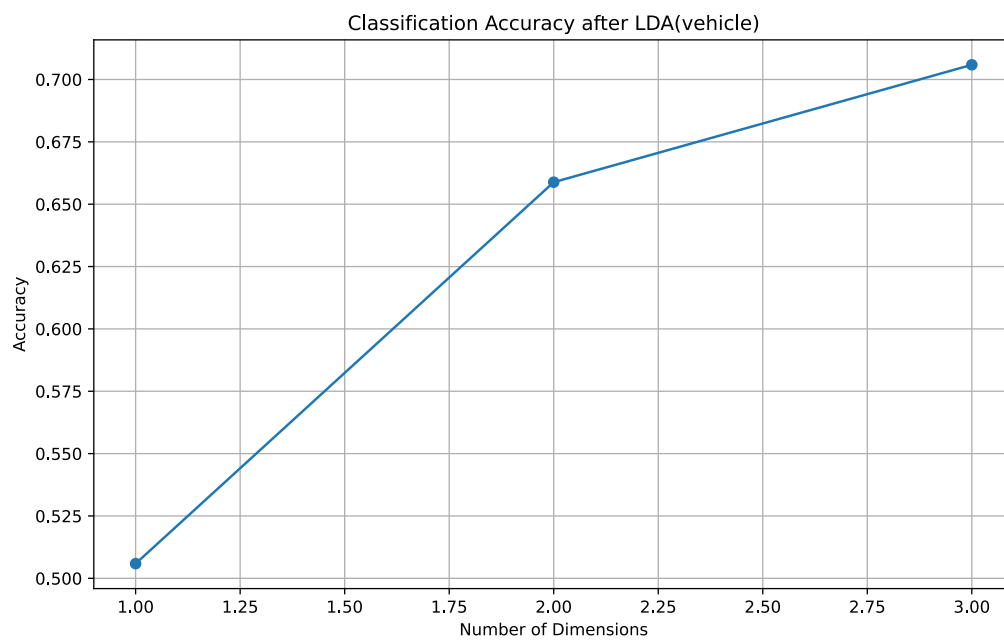
其中AT&T数据集结果如下图所示



降维维数设置为[10, 15, 20, 25, 30, 35]

可以看出, 当降维维数为30, 35时, 准确率最高, 为98.75%;

Vehicle数据集结果如下图所示



其中降维维数为[1, 2, 3]

降维维数为3时，准确率最高，为70.59%