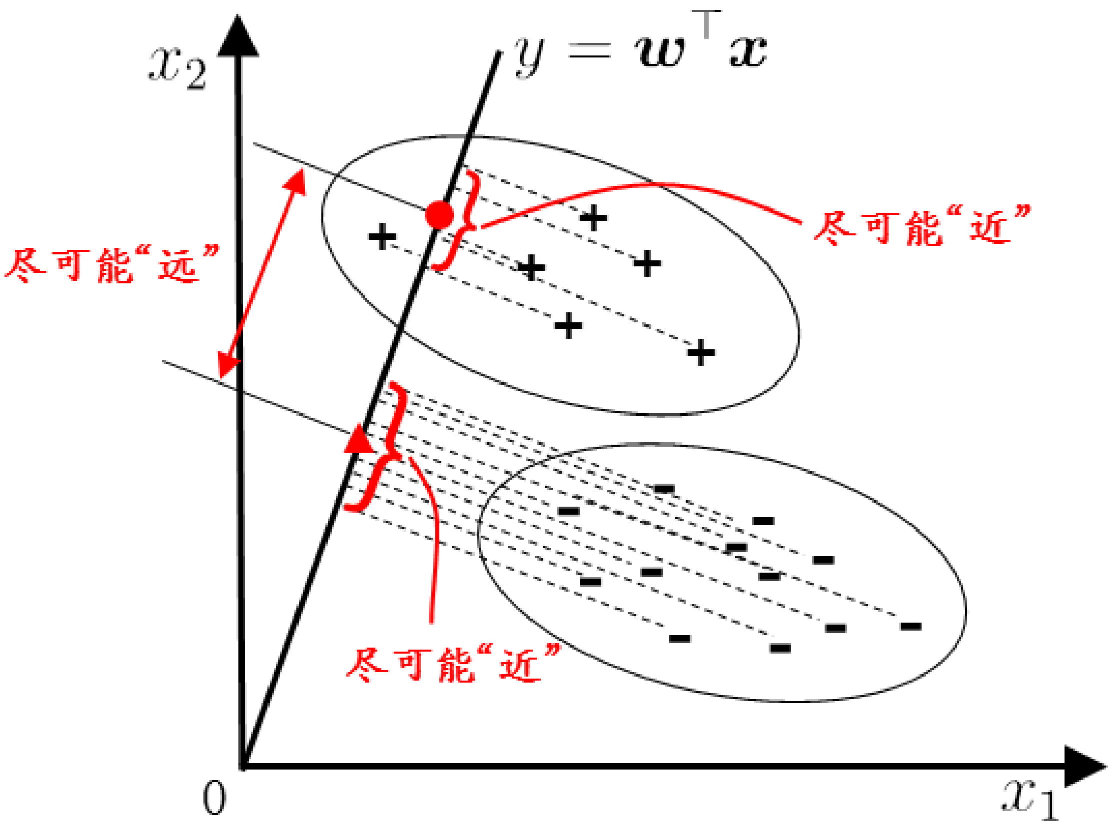
线性判别分析

Fisher判别分析

- 线性判别分析（Linear Discriminant Analysis） [Fisher, 1936]



LDA也可被视为一种监督降维技术

知识点之—— Fisher判别分析

建设单位： 大连理工大学
建 设 人： 刘倩

目录

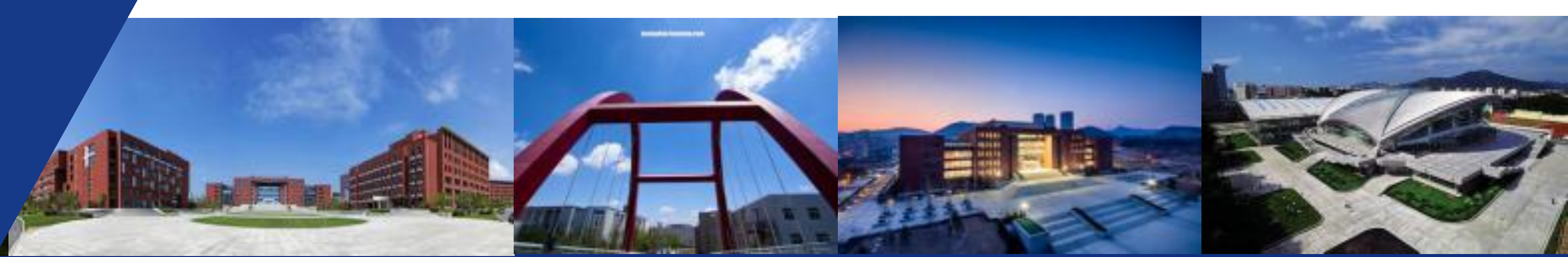
CONTENT

01 PCA vs Fisher判别分析

02 Fisher判别分析

03 总结与讨论

04 课外阅读材料推荐



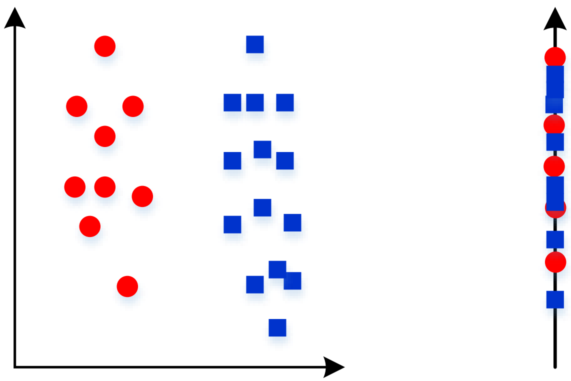
01

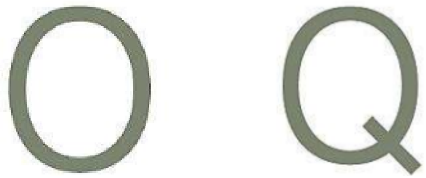
PART ONE

PCA vs Fisher判别分析

1 PCA vs Fisher判别分析

- **PCA将所有的样本作为一个整体对待，去寻找一个均方误差最小意义下的最优线性映射，而忽略了类别属性，即方差最大的方向不一定可用于分类**

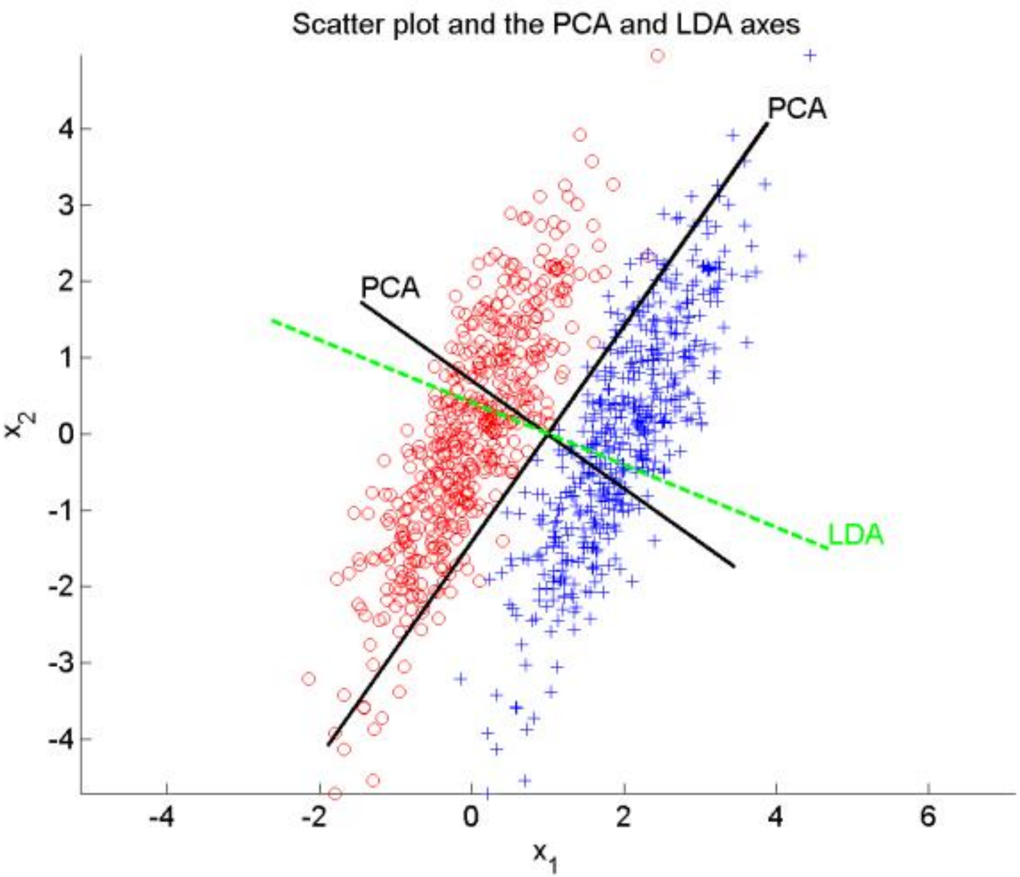


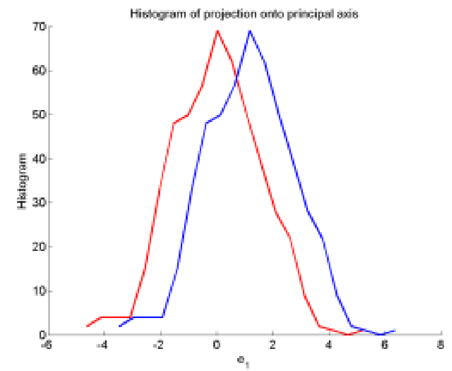


PCA能够识别字母O和Q的相似之处，但很可能把二者的那个“尾巴”特征抛弃

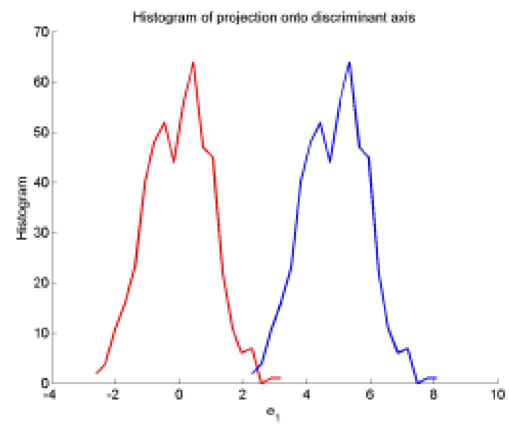
PCA	Fisher判别分析
寻找保留最多信息的主轴方向	寻找用来有效分类的方向

1 PCA vs Fisher判别分析/LDA





投影到PCA的主成分方向



投影到LDA的主方向

02

PART TWO

Fisher判别分析

2 Fisher判别分析

❑ 问题描述：

有N个d-维的样本$\mathbf{x}_1, \dots \mathbf{x}_N \in \mathbb{R}^d$, 分属于两个不同的类别$\omega_1$和$\omega_2$，对应的样本子集为$\mathcal{D}_1$和$\mathcal{D}_2$，其中$|\mathcal{D}_1| = N_1$，$|\mathcal{D}_2| = N_2$。

目标: 确定一条直线y（方向为$\mathbf{w}$），使$\mathbf{x}_1, \dots \mathbf{x}_N \in \mathbb{R}^d$投影到此直线的数据点$\mathrm{y}_1, \dots \mathrm{y}_N$（分属于$\mathcal{Y}_1$和$\mathcal{Y}_2$投影点子集）显著可分，即在投影 $\mathrm{y} = \mathbf{w}^T\mathbf{x}$ 下，**最大化**准则函数

$$J(\mathbf{w}) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

$|\tilde{m}_1 - \tilde{m}_2|$ 为投影后点的样本均值之差；

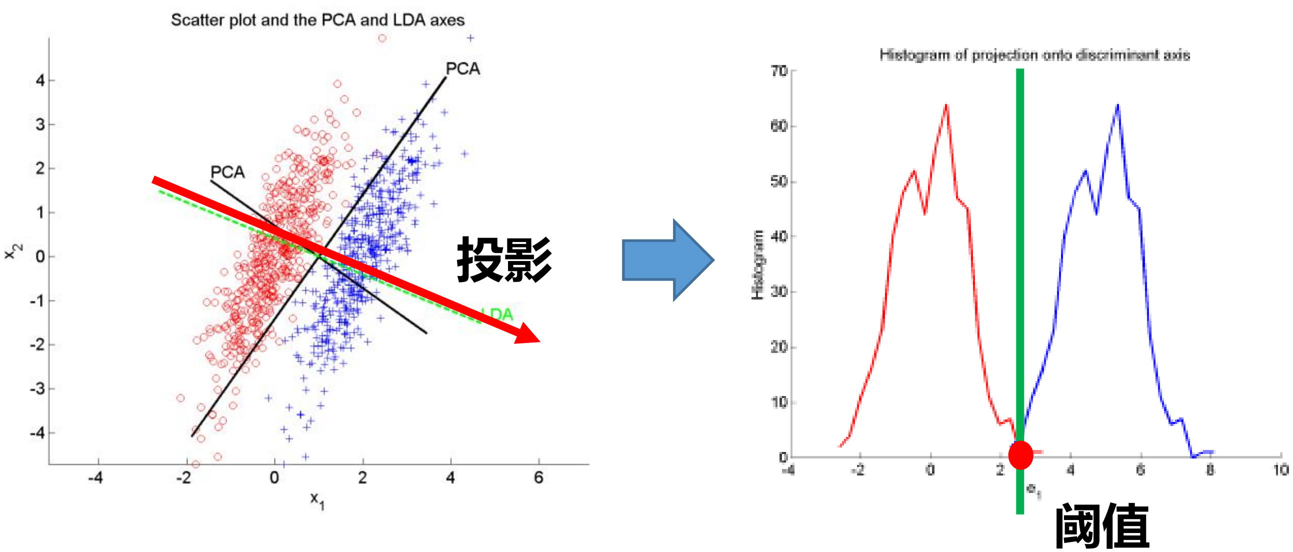
$\tilde{s}_1^2 + \tilde{s}_2^2$ 为投影后样本的总类内散布；

$\frac{1}{N}(\tilde{s}_1^2 + \tilde{s}_2^2)$ 为投影后数据总体方差估计.

2 Fisher判别分析

$$J(\mathbf{w})=\frac{|\tilde{m}_1-\tilde{m}_2|^2}{\tilde{s}_1^2+\tilde{s}_2^2}$$

- 任务1：**求解最佳投影w**使均值之间的差异相对于方差尽可能大。



- 任务2：**求解阈值**，即一维空间中将两类分开的点的位置。

2 Fisher判别分析

- **任务1：求解最佳投影w使均值之间的差异相对于方差尽可能大。**

$$J(\mathbf{w}) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

目标：将$J(\cdot)$写成w的表达式，求解使$J(\mathbf{w})$最大化的w

- 定义$\tilde{m}_i$为类别ω_i投影后的样本均值，$\mathbf{m}_i$为类别ω_i的样本均值，则

$$\tilde{m}_i = \frac{1}{n_i}\sum_{y\in\omega_i} y = \frac{1}{n_i}\sum_{\mathbf{x}\in D_i} \mathbf{w}^T\mathbf{x} = \mathbf{w}^T\frac{1}{n_i}\sum_{\mathbf{x}\in D_i}\mathbf{x} = \mathbf{w}^T\mathbf{m}_i$$

$$\boxed{\tilde{m}_i = \mathbf{w}^T\mathbf{m}_i}$$

- 定义$\mathbf{S}_\mathrm{B}$为总类间散布矩阵，则$\mathbf{S}_\mathrm{B} = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$，则

$$|\tilde{m}_1 - \tilde{m}_2|^2 = (\mathbf{w}^T\mathbf{m}_1 - \mathbf{w}^T\mathbf{m}_2)^2 = \mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T\mathbf{w}$$

$$\boxed{|\tilde{m}_1 - \tilde{m}_2|^2 = \mathbf{w}^T\,\mathbf{S}_\mathrm{B}\mathbf{w}}$$

2 Fisher判别分析

- 任务1：**求解最佳投影w使均值之间的差异相对于方差尽可能大。**

$$J(\mathbf{w})=\frac{|\tilde{m}_1-\tilde{m}_2|^2}{\tilde{s}_1^2+\tilde{s}_2^2}$$

目标：将$J(\cdot)$写成w的表达式，求解使$J(\mathbf{w})$最大化的w

- 定义$\mathbf{S}_i$为类内散布矩阵，则 $\mathbf{S}_i=\sum_{\mathbf{x}\in D_i}(\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^T$
- 定义$\mathbf{S}_W$为总类内散布矩阵，则 $\mathbf{S}_W=\mathbf{S}_1+\mathbf{S}_2$；
- 定义$\tilde{s}_i^2$为投影后的类内散布，则

$$\tilde{s}_i^2=\sum_{y\epsilon\omega_i}(y-\tilde{m}_i)^2=\sum_{\mathbf{x}\in D_i}(\mathbf{w}^T\mathbf{x}-\mathbf{w}^T\mathbf{m}_i)^2=\sum_{\mathbf{x}\in D_i}\mathbf{w}^T(\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^T\mathbf{w}$$

$$\boxed{\tilde{s}_i^2=\mathbf{w}^T\mathbf{S}_i\,\mathbf{w}}$$

$$\boxed{\tilde{s}_1^2+\tilde{s}_2^2=\mathbf{w}^T\mathbf{S}_W\,\mathbf{w}}$$

2 Fisher判别分析

- 任务1：**求解最佳投影w**使均值之间的差异相对于方差尽可能大。

$$J(\mathbf{w})=\frac{|\tilde{m}_1-\tilde{m}_2|^2}{\tilde{s}_1^2+\tilde{s}_2^2}$$

目标：将$J(\cdot)$写成w的表达式，求解使$J(\mathbf{w})$最大化的w

$$J(\mathbf{w})=\frac{|\tilde{m}_1-\tilde{m}_2|^2}{\tilde{s}_1^2+\tilde{s}_2^2}=\frac{\mathbf{w}^T\,\mathbf{S}_\mathrm{B}\mathbf{w}}{\mathbf{w}^T\mathbf{S}_\mathrm{W}\,\mathbf{w}}$$

这个目标函数在数学物理中常用，使得$J(\mathbf{w})$最大化的w应满足：

$$\mathbf{S}_\mathrm{B}\mathbf{w}=\lambda\mathbf{S}_\mathrm{W}\,\mathbf{w}$$

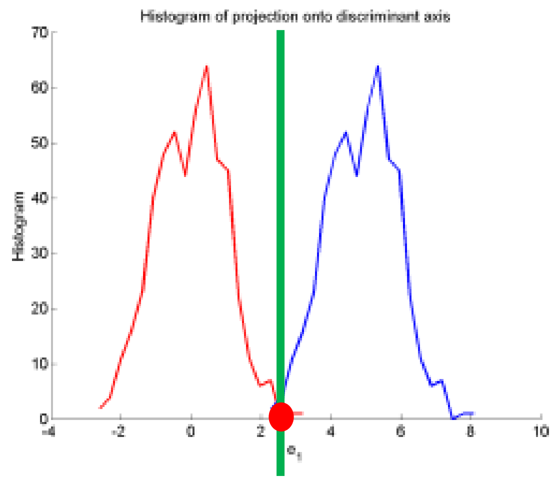
$\mathbf{S}_\mathrm{B}\mathbf{w}$ 总是位于$(\mathbf{m}_1-\mathbf{m}_2)$的方向上 → $\boxed{\mathbf{w}=\mathbf{S}_\mathrm{W}^{-1}(\mathbf{m}_1-\mathbf{m}_2)}$

2 Fisher判别分析

- **任务2：求解阈值y_0，即一维空间中将两类分开的点的位置。**

使用最佳边界判决条件：

$$\text{若 } \mathbf{y} = \mathbf{w}^{\mathrm{T}}\mathbf{x} < y_0\text{，则属于类别}\omega_1\text{；}$$
$$\text{反之，则属于类别}\omega_2\text{.}$$



实际操作：

1. 对投影后的数据进行平滑，或用一维高斯函数进行拟合；
2. 阈值y_0可以选择为使两个类的后验概率相同的y值。

大连理工大学 Dalian University of Technology

2 Fisher判别分析——多重判别分析

- **将Fisher判别准则从二类别问题推广到c类别问题**

  - **判别函数个数：** $c-\mathbf{1}$， $|\mathcal{D}_1| = N_1$， …， $|\mathcal{D}_c| = N_c$， $\mathbf{m}_i = \frac{1}{N_i}\sum_{\mathbf{x}\in\mathcal{D}_i}\mathbf{x}$
  - **投影问题：从d维空间向c -1维空间做投影（假设$d \geq c$）**
  - **总类内散布矩阵** $\mathbf{S}_\mathrm{W} = \sum_{i=1}^{c}\mathbf{S}_i$ 其中$\mathbf{m} = \frac{1}{N}\sum_{\mathbf{x}}\mathbf{x} = \frac{1}{N}\sum_{i=1}^{c}N_i\mathbf{m}_i$
  - **总类间散布矩阵** $\mathbf{S}_\mathrm{B} = \sum_{i=1}^{c}N_i(\mathbf{m}_i-\mathbf{m})(\mathbf{m}_i-\mathbf{m})^T$

$$J(\mathbf{W}) = \frac{|\mathbf{W}^T\mathbf{S}_\mathrm{B}\mathbf{W}|}{|\mathbf{W}^T\mathbf{S}_\mathrm{W}\mathbf{W}|}$$ **其中** $\mathbf{W} \in \mathbb{R}^{d\times(c-1)}$

使$J(\mathbf{W})$最大化的最优矩阵W的列向量是$\mathbf{S}_\mathrm{W}^{-1}\mathbf{S}_\mathrm{B}$的最大特征值对应的特征向量

$$\mathbf{S}_\mathrm{W}^{-1}\mathbf{S}_\mathrm{B}\mathbf{w}_i = \lambda_i\,\mathbf{w}_i$$

2 Fisher判别分析——具体步骤

1. 利用训练样本集合计算类内散布矩阵 S_W 和类间散度矩阵 S_B；

2. 计算 $S_W^{-1}S_B$ 的特征值；

3. 选择非0的c-1个特征值对应的特征矢量组成一个变换矩阵

   $W=[w_1, w_2, \ldots, w_{c-1}]$；

4. 训练和识别时，每一个输入的d维特征矢量x可以转换为c-1维的新特征矢量 $y=W^T x$

03

PART THREE

总结与讨论

3 总结与讨论

PCA	Fisher判别分析
寻找用于有效表示的主轴方向	寻找用来有效分类的方向

- 经Fisher判别分析变换后，新的坐标系不是一个正交坐标系
- 新的坐标维度最多为c-1，c为类别数
- 只有当样本数足够多时，才能够保证类内散度矩阵Sw为非奇异矩阵（存在逆阵），而样本数少时Sw可能是奇异矩阵

4 课外阅读材料推荐

❑ 特征降维——线性方法（PCA和Fisher判别分析相关）

- Eigenfaces for Recognition. M. Turk and A. Pentland
- Using Discriminant Eigenfeatures for Image Retrieval. D. Swets, J. Weng
- Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. P. Belhumeur, J. Hespanha, and D. Kriegman
- PCA versus LDA. A. Martinez, A. Kak

谢谢大家！