人脸民族特征抽取及识别

——基本原理与关键技术

肖雅夫 2015年6月5日

NOTICE

- 注意：本PPT讲解的LDA指线性判别分析（Linear Discriminant Analysis），非主题模型LDA（Latent Dirichlet Allocation）

Content

- 1. 论文简介
- 2. 民族人脸识别总体方案
- 3. 二元LDA线性判别分析
- 4. 多元LDA线性判别分析

1. 论文简介

1.1 论文基本信息

- 题目：人脸的民族特征抽取及其识别
- 作者：段晓东，王存睿，刘向东，刘慧

（大连民族学院非线性信息技术研究所，东北大学研究生院）

- 期刊：计算机科学，2010年8月
- 数据集：作者自建数据集，包含我国17个民族

1.2 论文贡献

- 论文解决了什么问题？

验证了代数特征和几何特征对于民族人脸识别的可行性

- 论文尚未解决哪些问题？

没有处理光照变化和姿态变化

1.3 论文涉及到的算法思想

- 涉及到的算法、思想或技术：

（1）线性判别分析LDA

（2）人脸弹性模板

（3）积分投影

（4）Gabor小波

（5）人脸拓扑结构

（6）决策树C5.0算法

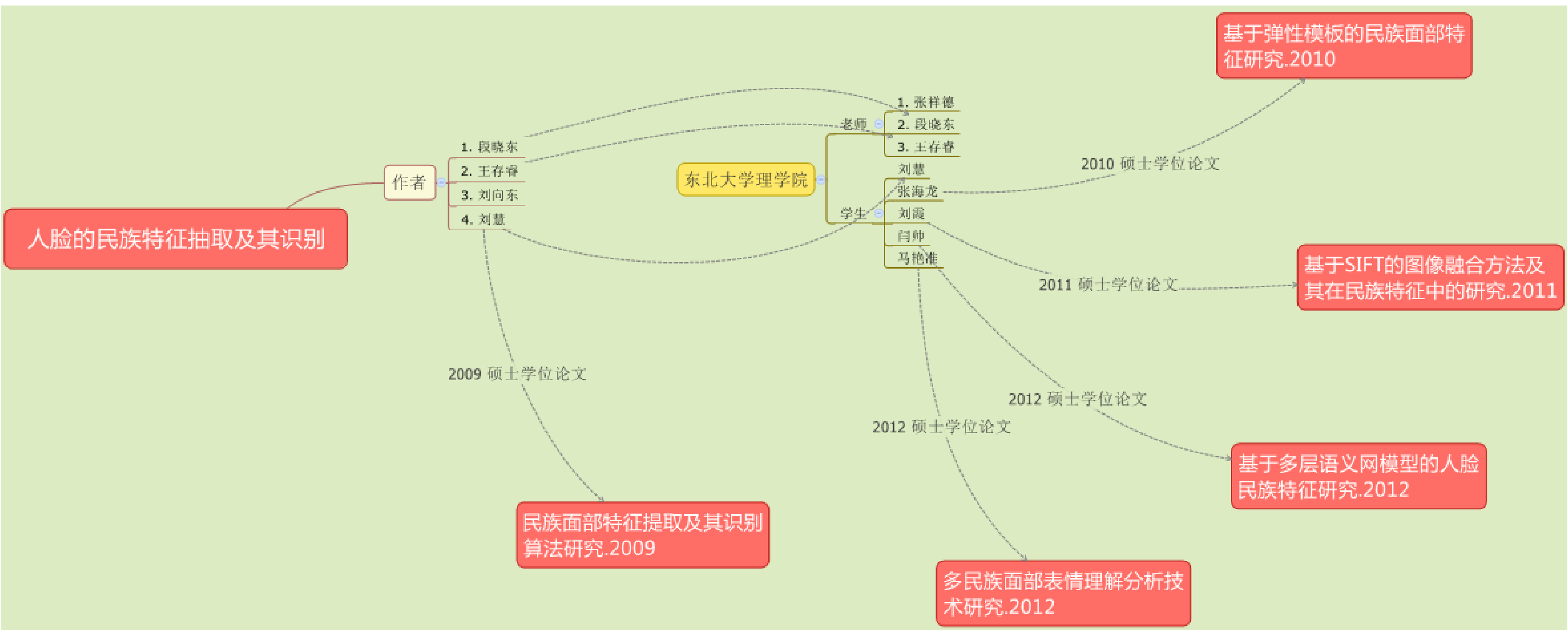
1.4 辅助参考文献

- 辅助参考文献：

张海龙.基于弹性模板的民族面部特征研究[D].东北大学.2010

刘慧.民族面部特征提取及其识别算法研究[D].东北大学.2009

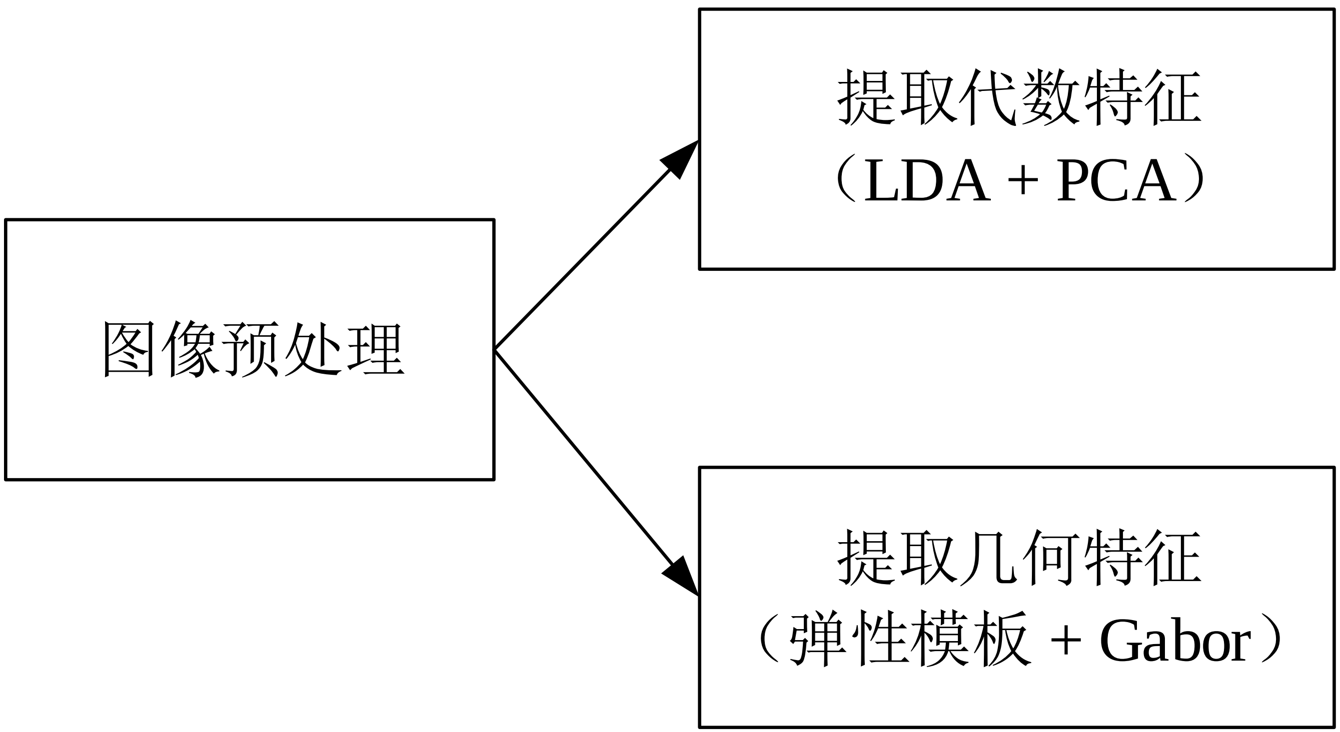
1.5 与作者相关的研究现状调查



2. 民族人脸识别总体方案

2. 民族人脸识别总体方案

• 本文做的主要工作：



2.1 图像预处理

- 1）图像预处理

图像预处理分别采用了灰度变换（灰度均衡化，去除背景）和几何变换（尺寸归一化，校正姿态）

注意：为便于采集人脸几何特征，本文将左右眼宽平均值作为归一化基向量，对图像进行等比例变换（原因：眼睛的眼宽差异相对方差最小。图像经过归一化之后，可以提高数据质量，相应地提高分类算法识别效能）

2.2 提取代数特征

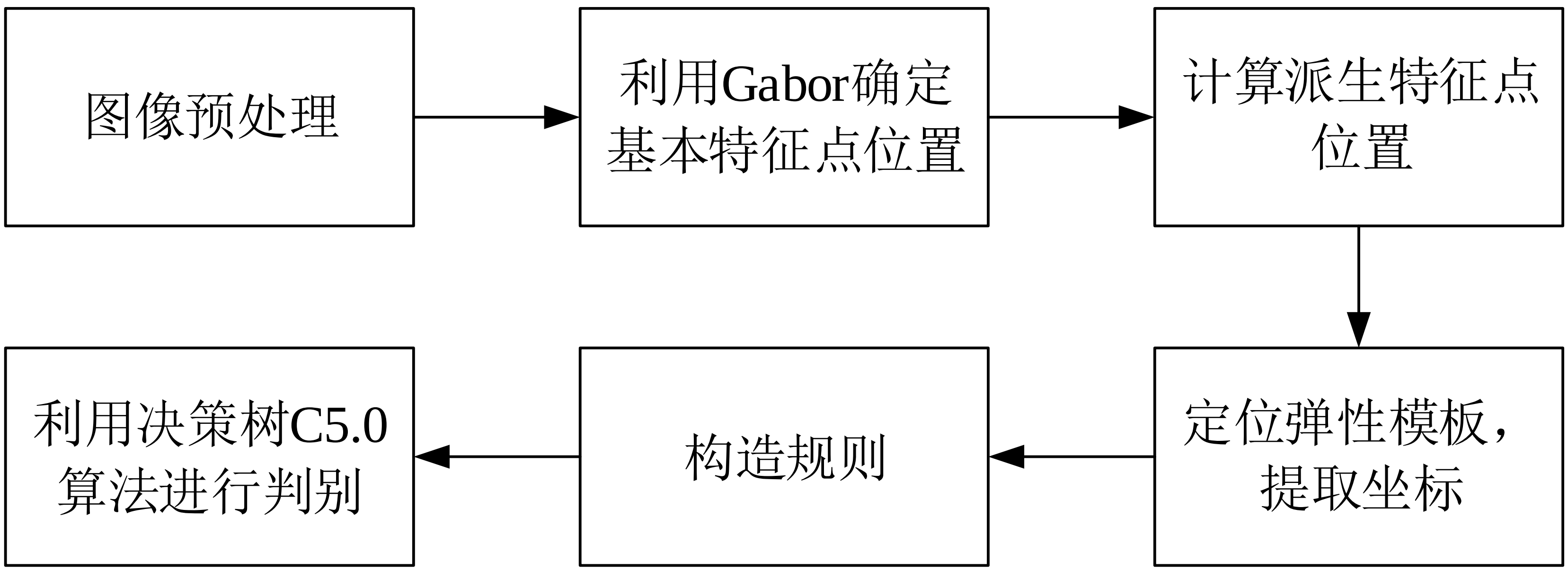
- 2）代数特征提取

LDA方法可从人脸图片信息选择出散布于类内的正交向量，并将其作为人脸代数特征的特征向量

LDA算法思路在下一章分析，这里只需知道它的作用就够了

2.3 几何特征提取与识别

- 几何特征算法总体思想



2.3 几何特征提取与识别

- 3）几何特征提取

通常包括面部主要特征点的位置以及指定点间的欧式距离或曲率、角度等

由这些特征便可以得到描述每张脸的特征矢量，将各民族人脸用一组几何特征矢量来表示

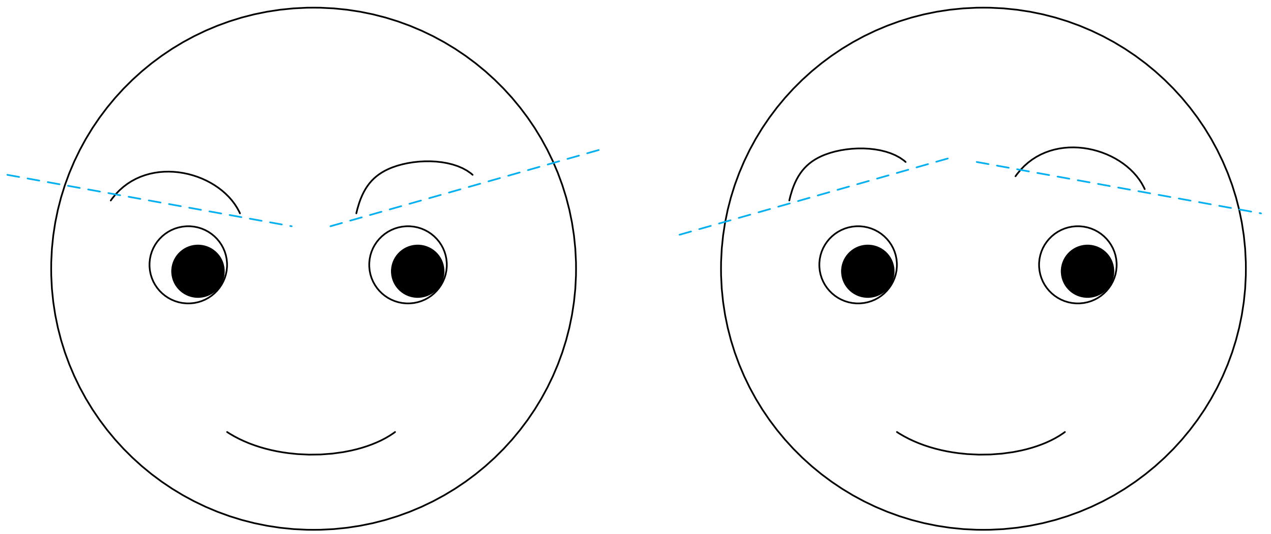
2.3 几何特征提取与识别

人类学相关研究表明，少数民族具有很多独特的面部特征，比如眉长轴和眼长轴及其之间夹角（眉长轴为眉毛最低点的联线，眼长轴为眼内外眦间的联线，例如蒙、藏族多为平行或外相交，维吾尔族多为平行或内相交）

同时，不同民族之间侧面人脸中的鼻根高度、位置和形状存在显著差异

2.3 几何特征提取与识别

比如下面的示意图，观察眉长轴，左边为维吾尔族人脸示意图，右边为蒙古族人脸示意图



2.3 几何特征提取与识别

- 下标描述了藏族、维吾尔族和壮族的人脸拓扑结构

民族	面部特征
藏族	**眼睛：**眼裂开度中等趋窄，多数人显示内眥褶 **鼻子：**鼻梁较直、细楞，鼻根高度大，鼻根凹小且靠近眉间 **嘴唇：**中等唇厚，唇形稍凸：上红唇多数为中等厚，下红唇比上红唇厚 **面型：**面部多呈长梭型及长梯型 **头发：**发质较硬；发黑直，间有波型及卷曲型 **耳朵：**１／３耳壳上显示达尔文结节，耳垂形状多为圆形，部分为方形，三角形甚少 **肤色：**黄褐、黑褐或古铜色

2.3 几何特征提取与识别

民族	面部特征
维吾尔族	**眼睛：**眼裂开度中等，眼为楔形结构，上睑皱褶明显，睑沿宽，睑裂大 **鼻子：**鼻根中等偏高，鼻尖高 **嘴唇：**上唇皮肤部高度多为中等和偏高，红唇多为中唇和唇 **面型：**卵圆形 **头发：**发黑间有黑褐色，头发有直、波、卷曲三型 **耳朵：**大多有达尔文节，耳垂多为圆形和方形 **肤色：**肤色呈黄、白或棕褐。

2.3 几何特征提取与识别

民族	面部特征
壮族	**眼睛：**眼裂开度较宽，蒙古褶较弱育，多数呈微显型 **鼻子：**鼻梁硬骨部稍下凹软骨部略上凸 **嘴唇：**上唇皮肤部一般为中等高度，明显前凸；红唇厚度数为中等偏厚 **面型：**中面型和阔面型为主 **头发：**发形直、发色黑 **耳朵：**耳壳上多数没有达尔文结节，耳垂向下，悬挂成舌状，即圆形耳垂

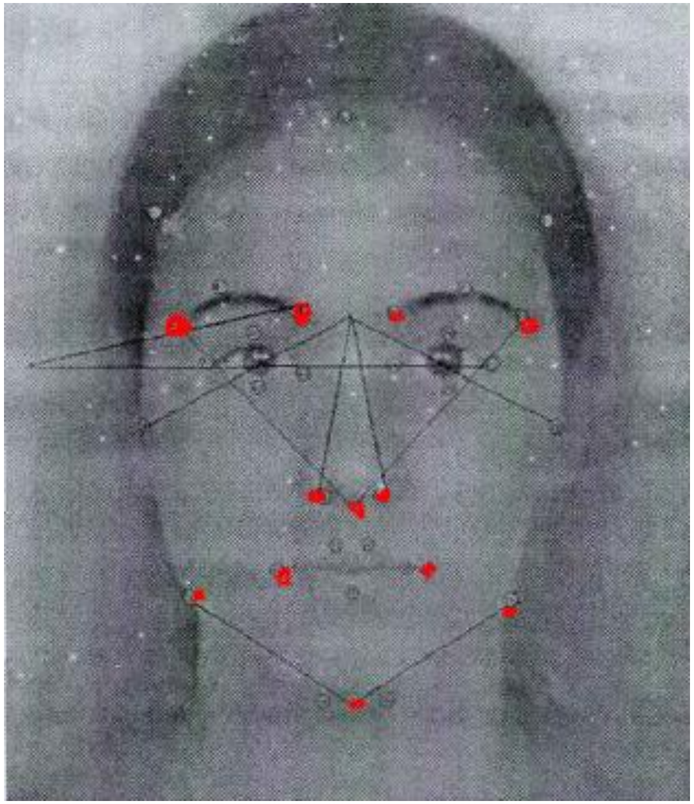
2.3 几何特征提取与识别

- 根据上述人脸拓扑结构，论文采用弹性模板和Gabor小波来提取几何特征。首先，什么是弹性模板？

- 弹性模板，就是在人脸上标注若干关键点，这些点具有各自的几何含义（比如点p_1表示左外眼角，点p_2表示左内眼角…），在不同的人脸上，这些点的具体像素坐标会有不同，看起来就像有弹性一样

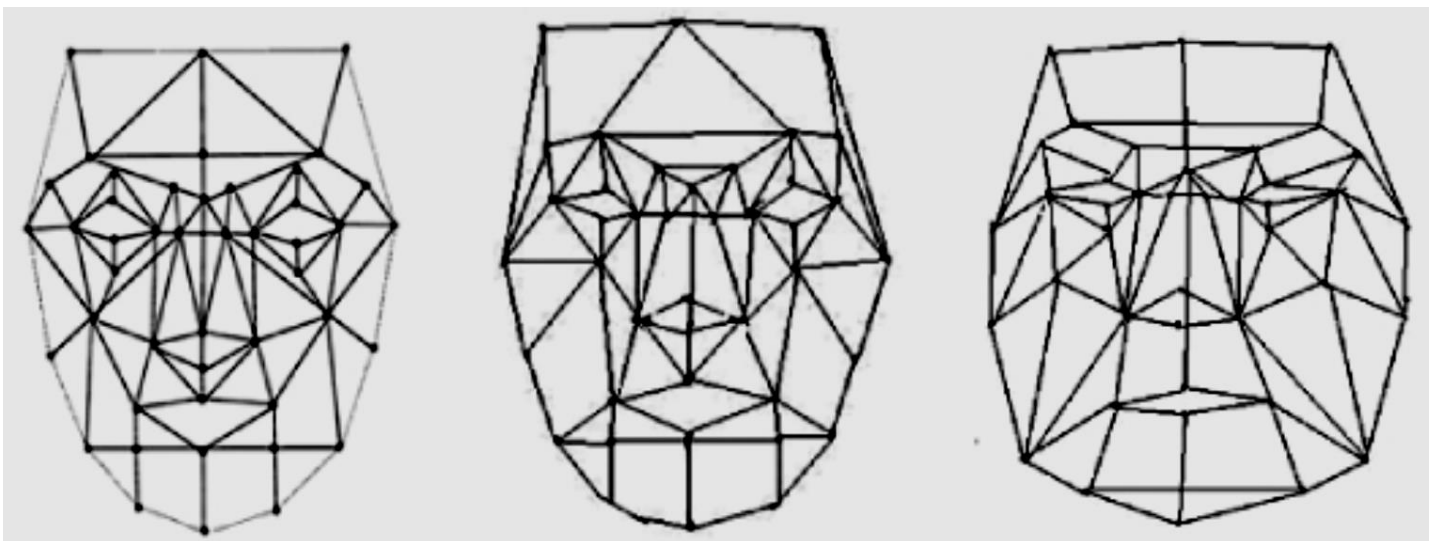
2.3 几何特征提取与识别

- 下图是一个几何模板的示例



2.3 几何特征提取与识别

- 论文采用的弹性模板示意图，左边为维吾尔族人脸，中间为藏族人脸，右边为蒙古族人脸



2.3 几何特征提取与识别

- 当然，也有研究者认为，弹性模板中的点并非都具有“弹性”，可能也存在一些点始终是固定的（张海龙.基于弹性模板的民族面部特征研究[D].东北大学.2010）

2.3 几何特征提取与识别

- 这里有一个问题：几何特征点的基础是各个关键点，那么这些关键点是如何提取的？
- 本文没有直接说明，但来自第三作者刘慧的硕士学位论文中给出了一种方法：积分投影法

2.3 几何特征提取与识别

- 积分投影法（Integral Projection Function, IPF）的基本思想是：对正面人脸图像进行水平积分投影，然后利用曲线的极值来定位各特征点
- 根据图像某些方向上的投影分布特征来进行检测，本质上是一种统计方法
- 积分投影法主要有水平投影和垂直投影

2.3 几何特征提取与识别

- 设$G(x,y)$表示图像(x,y)处的灰度值，在图像$[y_1,y_2]$和$[x_1,x_2]$区域的水平积分$M_h(x)$表示为：

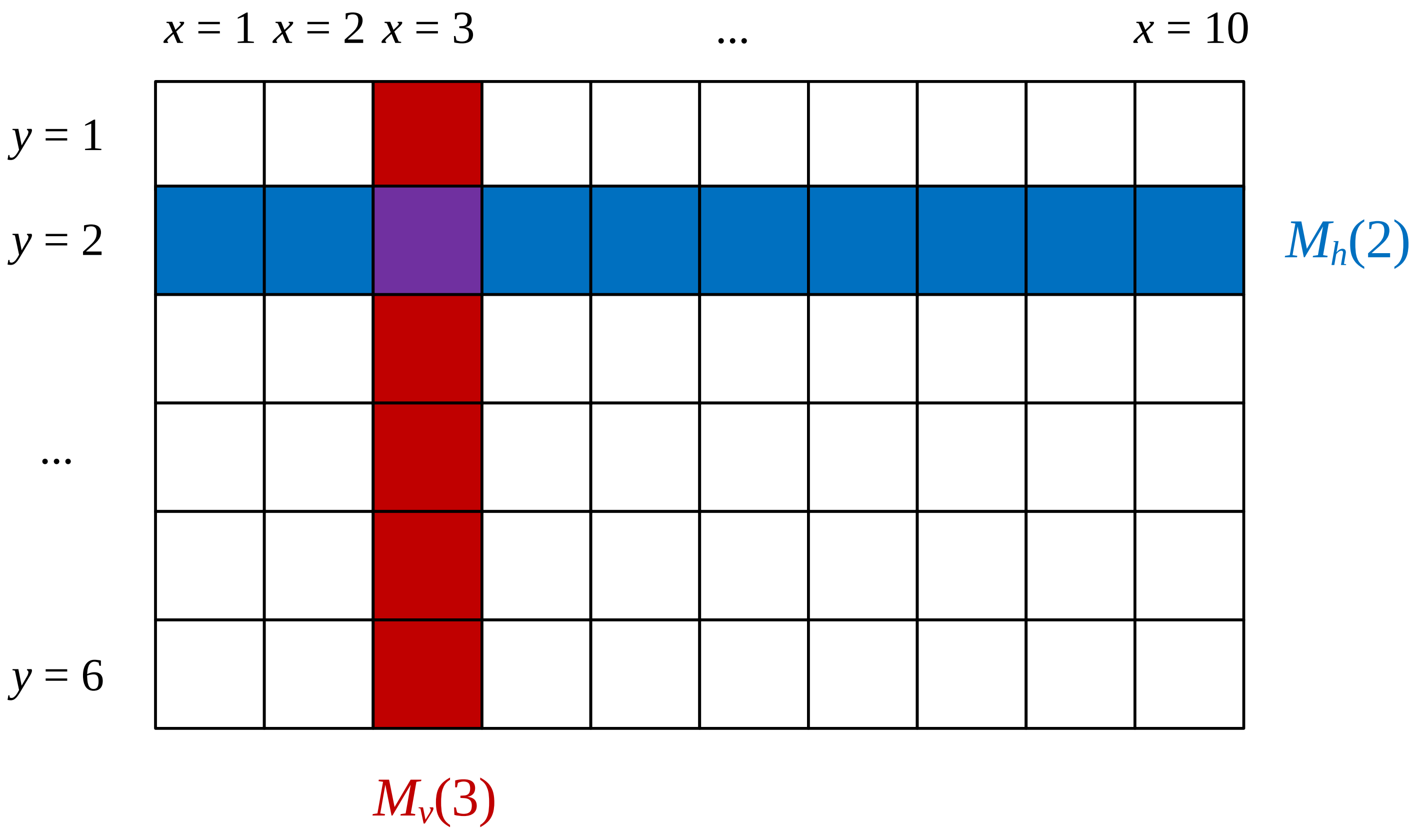
$$M_h(y)=\frac{1}{x_2-x_1}\sum_{x_1}^{x_2}G(x,y)$$

- 垂直积分表示为：

$$M_v(x)=\frac{1}{y_2-y_1}\sum_{y_1}^{y_2}G(x,y)$$

2.3 几何特征提取与识别

- 水平积分和垂直积分示意图



2.3 几何特征提取与识别

- 对每一行（或者每一列）提取了积分之后，可以得到相应的函数曲线，如下图所示：

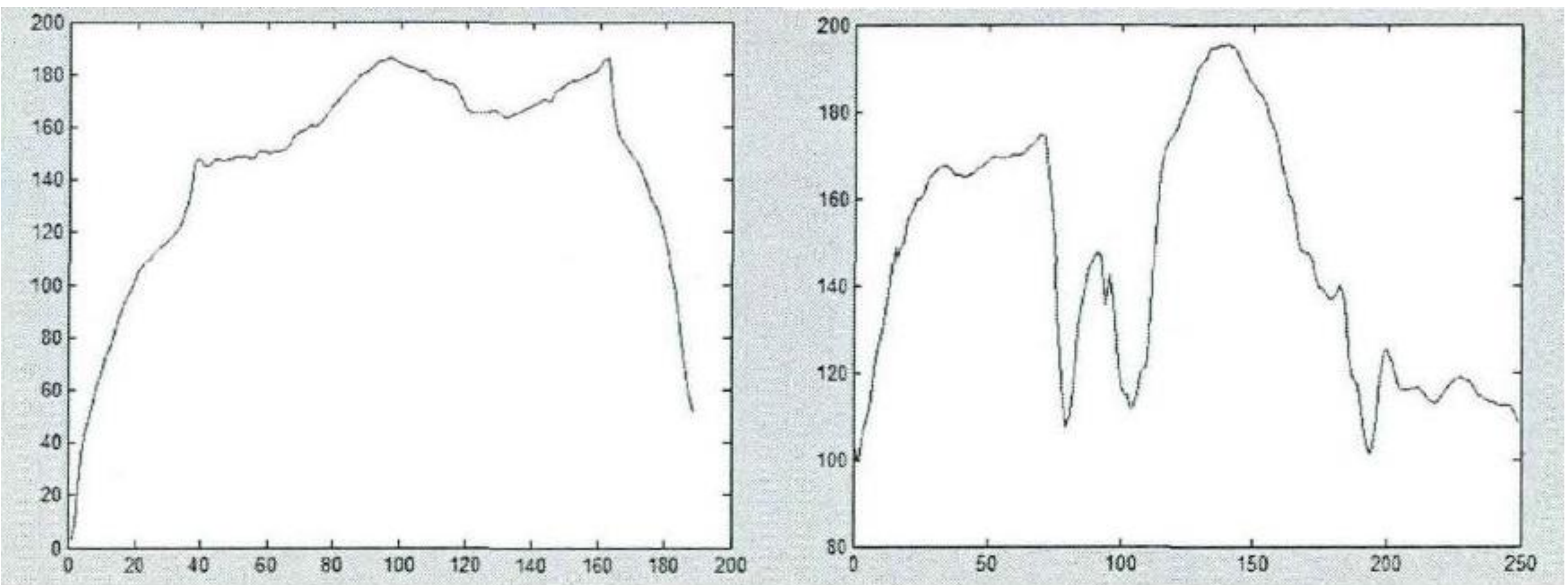


图 4.9 (a) 垂直投影曲线 (b) 水平投影曲线

2.3 几何特征提取与识别

- 得到了函数曲线之后，通过梯度计算来找到曲线的极值点（波峰和波谷），进而对特征点进行定位
- 很多个特征点计算出来之后，就得到了人脸弹性模板

2.3 几何特征提取与识别

- 另一个问题，已经能够得到弹性模板，为什么还需要Gabor小波？

- 它可以在弹性模板的基础上进行Gabor特征提取，形成人脸属性标号图，最后民族人脸识别的过程就转换为各民族人脸标号图之间的比较（援引文献《民族面部特征提取及其识别算法研究_刘慧》P49），关于标号图，后续会有详细介绍

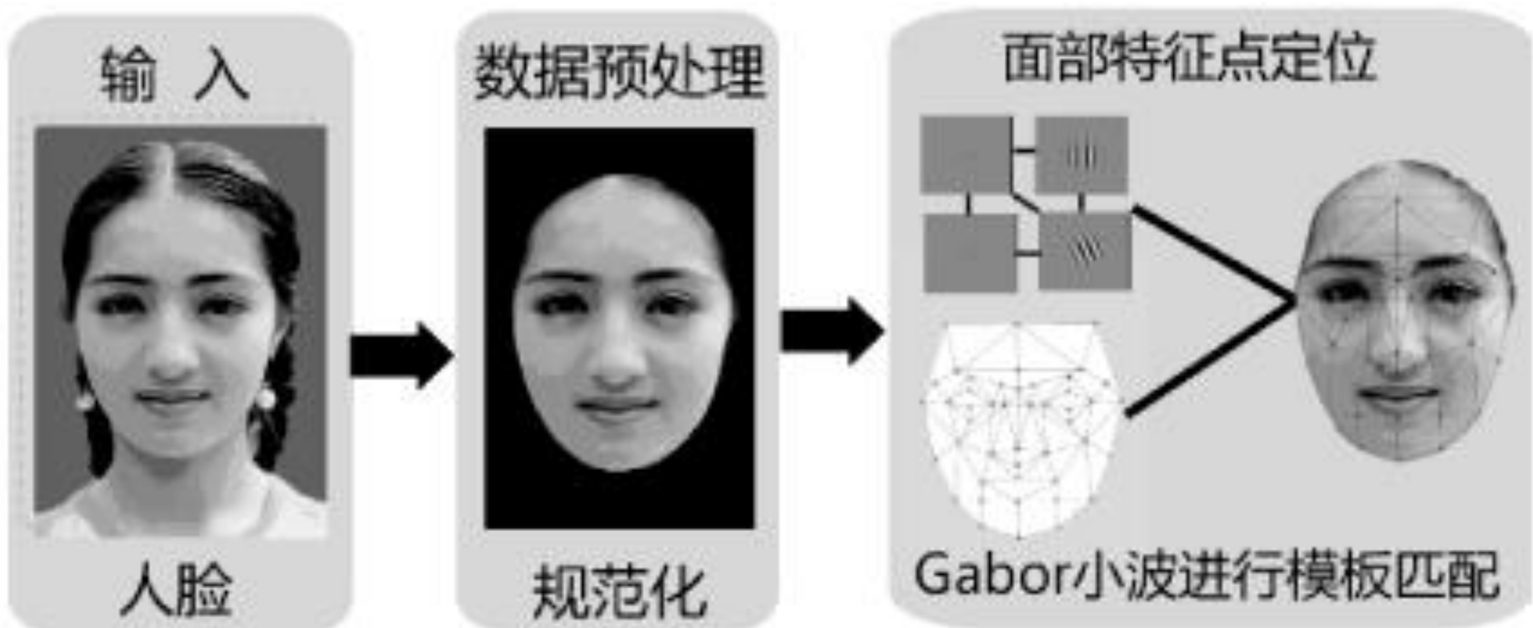
2.3 几何特征提取与识别

- 对于本文而言，利用Gabor小波提取图像不同区域的不同频率成分和方向的纹理信息

- 关于Gabor的详细信息会在PPT后续章节介绍，本节仅阐述上层业务逻辑

2.3 几何特征提取与识别

- 论文对Gabor的应用示意图：



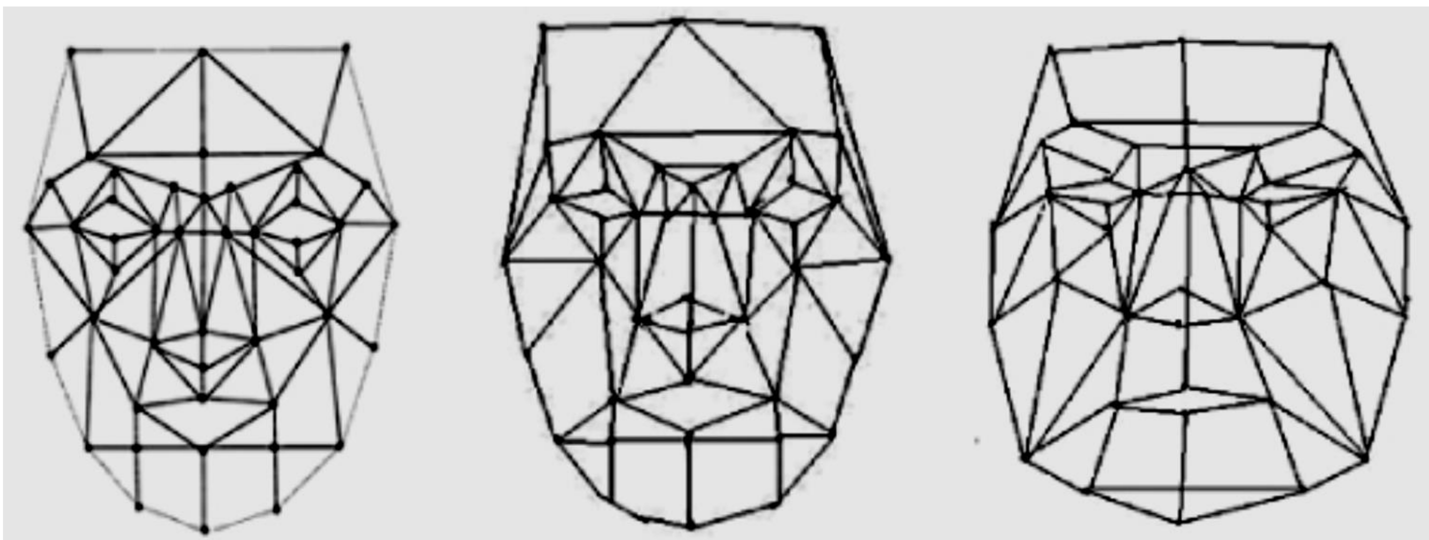
2.3 几何特征提取与识别

- 论文的原话：“本文建立人脸的特征点和弹性模板的对应部位的特征之间的联系，抽取能表达人脸各器官拓扑关系的几何特征”

- 论文的原话：“利用Gabor小波通过弹性束图匹配来确定图 3 所示基本特征点的位置、基本特征点的选取（如眉、眼睛、鼻子、嘴巴以及脸部轮廓等）。”（论文所述的图3见下页）

2.3 几何特征提取与识别

- 即：利用Gabor来定位特征点



2.3 几何特征提取与识别

- 定位了特征点后（包括基本特征点和派生特征点）

- 派生特征点是根据基本特征点算出来的，比如求中点等方法

2.3 几何特征提取与识别

- 得到了特征点后，就得到了一个弹性模板（每个点都用几何参数表示）
- 每一张脸都对应一组这样的参数，利用这些数据进行训练，可以得到分类规则

2.3 几何特征提取与识别

- 论文中展示的分类规则树截图：

规则 1　规则准确率 90％　鼻宽＜＝0.589[类别:藏族]＝＞藏族

鼻宽＞0.589[类别:维吾尔族]

嘴角到同侧内眼角＜＝2.030[类别:壮族]＝＞壮族

嘴角到同侧内眼角＞2.030[类别:维吾尔族]

鼻翼处脸宽＜＝4.462[类别:维吾尔族]

嘴宽＜＝1.481[类别:藏族]＝＞藏族

嘴宽＞1.481[类别:维吾尔族]＝＞维吾尔族

鼻翼处脸宽＞4.462[类别:壮族]＝＞壮族

规则 2　规则准确率 72.59％

下巴到嘴角＜＝1.526[类别:壮族]

下巴到鼻尖＜＝1.913[类别:藏族]＝＞藏族

下巴到鼻尖＞1.913[类别:壮族]＝＞壮族

下巴到嘴角＞1.526[类别:维吾尔族]

鼻翼处脸宽＜＝4.220[类别:维吾尔族]＝＞维吾尔族

鼻翼处脸宽＞4.220[类别:藏族]＝＞藏族

规则 3　规则准确率 87.18％

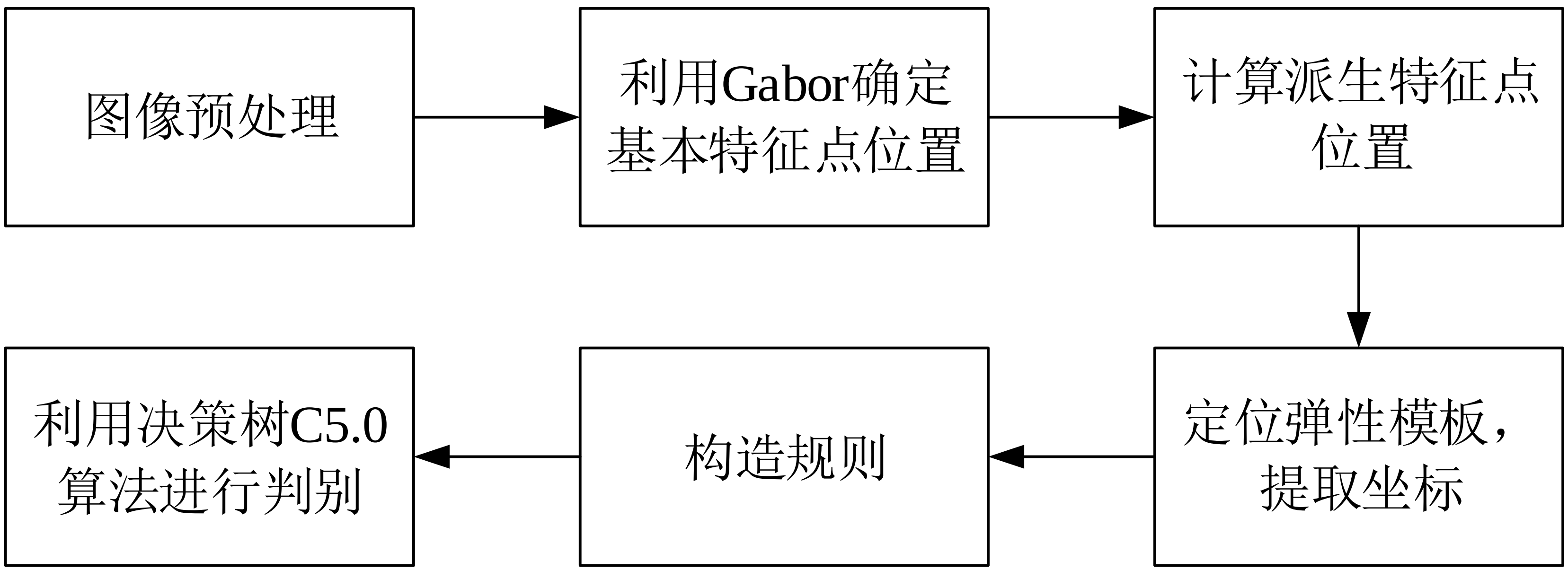
2.3 几何特征提取与识别

- 得到了规则树之后，就可以用这棵树对测试样本进行预测

	藏族	维吾尔族	壮族
藏族测试集	0.885714	0.071429	0.042857
维吾尔族测试集	0.071429	0.900000	0.028571
壮族测试集	0.042857	0.014286	0.942857

2.3 几何特征提取与识别

- 最后回顾一下几何特征抽取与识别的过程



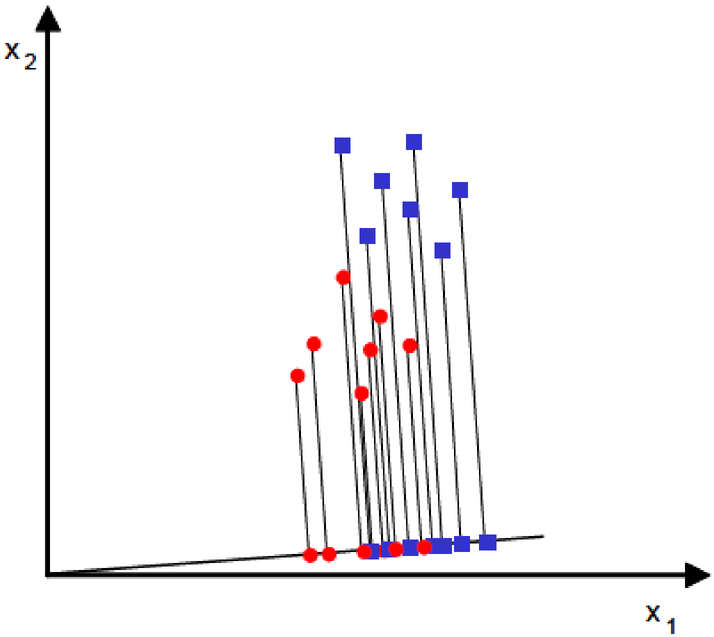
3. 二元LDA线性判别分析

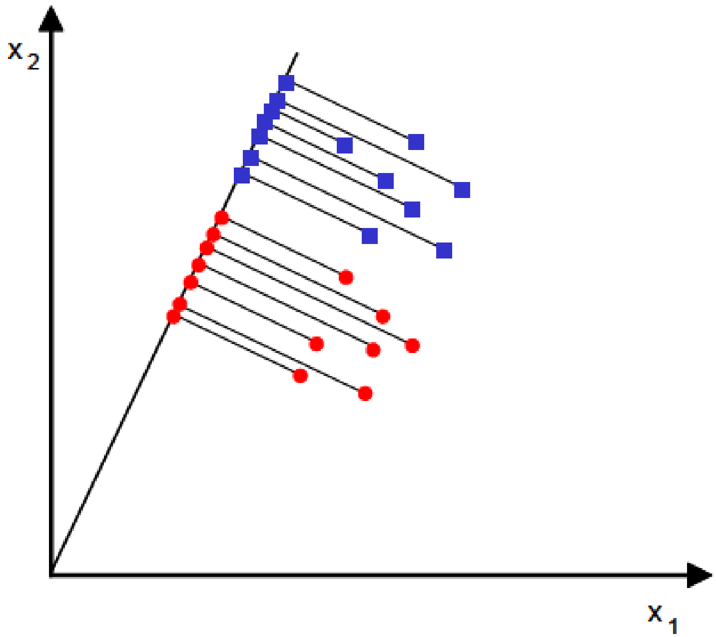
3.1 二元LDA问题引入

- 通常情况下，待匹配人脸要和人脸库内的多张人脸匹配，所以这是一个多分类的情况。出于简单考虑，可以先介绍二类的情况然后拓展到多类。

- 假设有二维平面上的两个点集$\{X_1\}$和$\{X_2\}$，它们的分布如下图所示（分别以蓝点和红点表示数据）：

3.1 二元LDA问题引入





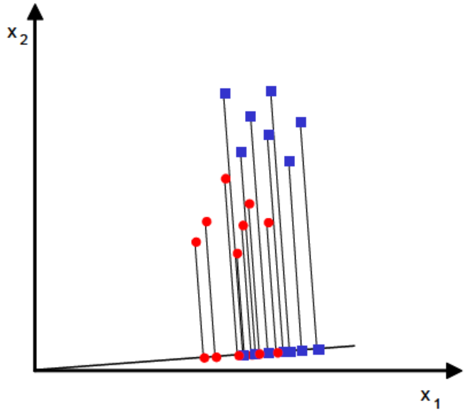
3.1 二元LDA问题引入

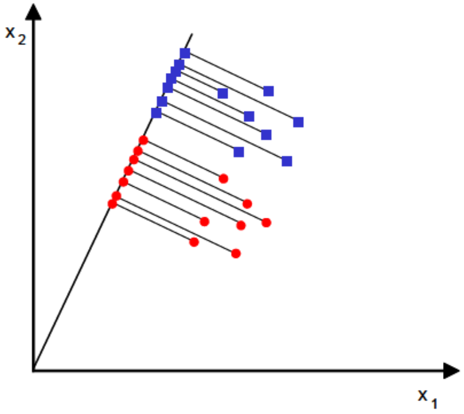
- 原有数据是散布在平面上的二维数据，如果想用一维的量（比如到圆点的距离）来合理的表示而且区分开这些数据，该怎么办呢？

- 一种有效的方法是找到一个合适的向量 $\boldsymbol{w}$ 进行投影，LDA算法约定这个向量是经过原点的

3.1 二元LDA问题引入

- 看刚才这张图片，我们假设有两种投影方案（即有两个$\boldsymbol{w}$，两者都是过原点的直线），似乎右边的方案更好一些
- Question：如何让计算机找到这样一个最优的投影向量？





3.1 二元LDA问题引入

- 首先定义问题：

给定N个样本，它们都是d维：

$$x^{(i)}\left\{x_1^{(i)},x_2^{(i)},\cdots,x_d^{(i)}\right\}$$

其中有N_1个样本属于类别ω_1，另外N_2个样本属于类别ω_2

3.1 二元LDA问题引入

- 通俗地讲，现在我们觉得原始特征数太多，想将d维特征降到只有一维，而又要保证类别能够“清晰”地反映在低维数据上，也就是这一维就能决定每个样例的类别

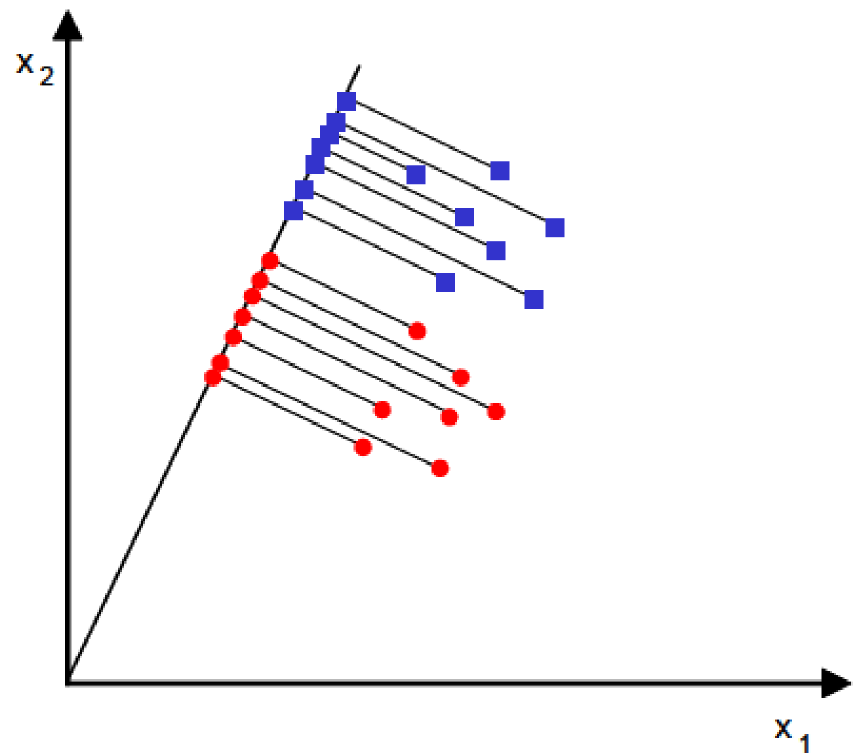
- 维数降到只有一维，那会是什么？它代表一种距离

3.1 二元LDA问题引入

- 假设我们找到的这个最优向量为$\boldsymbol{w}$，那么样例x（d维）到$\boldsymbol{w}$上的投影可以用下式来计算

$$y = \boldsymbol{w}^T x \quad (1)$$

这里的y就是x投影到直线上的点到原点的距离，参考右图



3.2 投影

- 首先计算每一类训练样本的均值（即中心点）

$$\mu_i = \frac{1}{N_i}\sum_{x\in\omega_i} x \quad (2)$$

- 样本投影到$\boldsymbol{w}$上的均值为

$$\widetilde{\mu}_i = \frac{1}{N_i}\sum_{y\in\omega_i} y = \frac{1}{N_i}\sum_{y\in\omega_i} \boldsymbol{w}^T x = \boldsymbol{w}^T \mu_i \quad (3)$$

3.3 定义求解标准

- 最佳投影向量$\boldsymbol{w}$的标准：1、不同的分类得到的投影点要尽量分开；2、同一个分类投影后得到的点要尽量聚合

- 据此定义$|\widetilde{\mu_1}-\widetilde{\mu_2}|$，它表示不同类别投影中心的距离

$$|\widetilde{\mu_1}-\widetilde{\mu_2}|=|\boldsymbol{w}^T(\mu_1-\mu_2)| \quad (4)$$

显然，$|\widetilde{\mu_1}-\widetilde{\mu_2}|$越大越好

3.3 定义求解标准

- 又定义投影后的散列值（scatter）$\widetilde{s_i}^2$，即对投影后的类求散列值

$$\widetilde{s_i}^2 = \sum_{y \in \omega_i} (y - \widetilde{\mu_i}) \quad (5)$$

跟我们熟悉的方差相比，散列值只是没有除以样本数量。它的几何意义是样本点的密集程度，值越大，越分散，反之，越集中。

3.3 定义求解标准

分子：两类中心距离

- 由上述(4)式和(5)式，得到最终希望的求解标准度量公式：

$$J(\boldsymbol{w}) = \frac{|\widetilde{\mu_1} - \widetilde{\mu_2}|^2}{\widetilde{s_1}^2 + \widetilde{s_2}^2} \quad (6)$$

分母：每个类自己的散列度加和

即

$$\max_{\boldsymbol{w}} J(\boldsymbol{w}) = \max_{\boldsymbol{w}} \frac{|\widetilde{\mu_1} - \widetilde{\mu_2}|^2}{\widetilde{s_1}^2 + \widetilde{s_2}^2} \quad (7)$$

3.3 定义求解标准

- 以$\widetilde{s_1}^2$为例，将(6)式中的$\widetilde{s_1}^2$展开，得

$$\widetilde{s_1}^2 = \sum_{y \in \omega_i} (y - \widetilde{\mu_1})^2 = \sum_{y \in \omega_i} (\boldsymbol{w}^T x - \boldsymbol{w}^T \mu_1)^2$$

$$= \sum_{y \in \omega_i} \boldsymbol{w}^T (x - \mu_1)(x - \mu_1)^T \boldsymbol{w}$$

(8)

3.3 定义求解标准

- 同理，得到：

$$\begin{cases} \widetilde{s_1}^2 = \sum_{y\in\omega_1} \boldsymbol{w}^T(x-\mu_1)(x-\mu_1)^T\,\boldsymbol{w} \quad (9) \\ \widetilde{s_2}^2 = \sum_{y\in\omega_2} \boldsymbol{w}^T(x-\mu_2)(x-\mu_2)^T\,\boldsymbol{w} \quad (10) \end{cases}$$

3.3 定义求解标准

- 即

$$\widetilde{s}_i^{\,2} = \sum_{y\in\omega_i} \boldsymbol{w}^T (x-\mu_i)(x-\mu_i)^T\,\boldsymbol{w}$$

$$= \boldsymbol{w}^T \sum_{y\in\omega_i} (x-\mu_i)(x-\mu_i)^T\,\boldsymbol{w}$$

(11)

3.3 定义求解标准

- 在上述(11)式中，令中间红颜色的部分为

$$\sum_{y\in\omega_i}(x-\mu_i)(x-\mu_i)^T = S_i \quad (12)$$

跟协方差矩阵相比，它仅仅没有除以样本数目，这样的矩阵S_i称为散列矩阵

3.3 定义求解标准

- 定义类内散列矩阵（Within-class scatter matrix）S_w

$$S_W = S_1 + S_2$$

$$= \sum_{y \in \omega_1} (x - \mu_1)(x - \mu_1)^T + \sum_{y \in \omega_2} (x - \mu_2)(x - \mu_2)^T \tag{13}$$

3.3 定义求解标准

- 于是，上述(11)式可以变为

$$\tilde{s}_i^{\,2} = \boldsymbol{w}^T S_i \boldsymbol{w} \quad (14)$$

- 用替换法，还可以得到

$$\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2} = \boldsymbol{w}^T S_1 \boldsymbol{w} + \boldsymbol{w}^T S_2 \boldsymbol{w} = \boldsymbol{w}^T (S_1 \boldsymbol{w} + S_2 \boldsymbol{w})$$

$$= \boldsymbol{w}^T [(S_1 + S_1)\boldsymbol{w}] = \boldsymbol{w}^T (S_1 + S_1)\boldsymbol{w} = \boldsymbol{w}^T S_W \boldsymbol{w} \quad (15)$$

3.3 定义求解标准

- 这样，(6)式$J(\boldsymbol{w})=\frac{|\widetilde{\mu_1}-\widetilde{\mu_2}|^2}{\widetilde{s_1}^2+\widetilde{s_2}^2}$中的分母部分就进行了一次转换，下面观察分子

- 对分子进行展开，有

$$|\widetilde{\mu_1}-\widetilde{\mu_2}|^2=(\widetilde{\mu_1}-\widetilde{\mu_2})^2=(\boldsymbol{w}^T\mu_1-\boldsymbol{w}^T\mu_2)^2$$
$$=\boldsymbol{w}^T\underbrace{(\mu_1-\mu_2)(\mu_1-\mu_2)^T}\boldsymbol{w}=\boldsymbol{w}^T S_B\boldsymbol{w}\quad(16)$$

S_B即类间散列矩阵（Between-class scatter matrix）

3.3 定义求解标准

- (16)式中产生了一个小式子(16.1)

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \quad (16.1)$$

3.3 定义求解标准

- 关于(16)式的推导

$$
\begin{aligned}
&\boldsymbol{w}^T(\mu_1-\mu_2)(\mu_1-\mu_2)^T\boldsymbol{w} \\
&=(\boldsymbol{w}^T\mu_1-\boldsymbol{w}^T\mu_2)(\mu_1^T\boldsymbol{w}-\mu_2^T\boldsymbol{w}) \\
&=(\boldsymbol{w}^T\mu_1-\boldsymbol{w}^T\mu_2)(\boldsymbol{w}^T\mu_1-\boldsymbol{w}^T\mu_2)^T \\
&=(\boldsymbol{w}^T\mu_1-\boldsymbol{w}^T\mu_2)^2
\end{aligned}
$$

其中，对于矩阵A，A^2视为AA^T

结论1：

$$\boldsymbol{w}^T(\mu_1-\mu_2)(\mu_1-\mu_2)^T\boldsymbol{w}=(\boldsymbol{w}^T\mu_1-\boldsymbol{w}^T\mu_2)^2$$

3.3 定义求解标准

类间散列矩阵

- 这样，(6)式可以表示为

$$J(\boldsymbol{w}) = \frac{|\widetilde{\mu_1} - \widetilde{\mu_2}|^2}{\widetilde{s_1}^2 + \widetilde{s_2}^2} = \frac{\boldsymbol{w}^T S_B \boldsymbol{w}}{\boldsymbol{w}^T S_W \boldsymbol{w}} \quad (17)$$

类内散列矩阵

对(17)式的分母进行归一化，得到约束条件

$$\|\boldsymbol{w}^T S_W \boldsymbol{w}\| = 1 \quad (18)$$

3.4 拉格朗日

- 利用$J(\boldsymbol{w})$以及约束条件构造拉格朗日表达式

$$c(w) = \boldsymbol{w}^T S_B \boldsymbol{w} - \lambda(\boldsymbol{w}^T S_W \boldsymbol{w} - 1) \quad (19)$$

- 对$\boldsymbol{w}$求偏导，并令其为零，得

$$\frac{\partial c}{\partial \boldsymbol{w}} = 2S_B\boldsymbol{w} - 2\lambda S_W \boldsymbol{w} = 0 \Rightarrow S_B \boldsymbol{w} = \lambda S_W \boldsymbol{w} \quad (20)$$

这里面用到了矩阵微积分，求导时可以简单把$\boldsymbol{w}^T S_W \boldsymbol{w}$当作$S_W \boldsymbol{w}^2$来看待

3.4 拉格朗日

- 观察(20)式，$S_B\boldsymbol{w} = \lambda S_W\boldsymbol{w}$，如果$S_W$可逆，那么可以在求导结果两边都乘以$S_W^{-1}$，得到

$$S_W^{-1}S_B\boldsymbol{w} = \lambda S_W^{-1}S_W\boldsymbol{w} \quad (21)$$

进而

$$S_W^{-1}S_B\boldsymbol{w} = \lambda\boldsymbol{w} \quad (22)$$

不难看出，此时的$\boldsymbol{w}$就是特征向量，λ就是对应的特征值

3.4 拉格朗日

- 在式(22)中，S_W^{-1}和S_B均可以求出（见(13)式和(16)式），因此特征向量$\boldsymbol{w}$和对应的特征值λ也能够求出

- 附：回顾(13)式和(16)式

$$S_W = S_1 + S_2 \quad (13)$$

$$|\widetilde{\mu_1} - \widetilde{\mu_2}|^2 = (\widetilde{\mu_1} - \widetilde{\mu_2})^2 = (\boldsymbol{w}^T\mu_1 - \boldsymbol{w}^T\mu_2)^2$$

$$= \boldsymbol{w}^T(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T\boldsymbol{w} = \boldsymbol{w}^T S_B \boldsymbol{w} \quad (16)$$

3.4 拉格朗日

- 上述公式(22) $S_W^{-1} S_B \boldsymbol{w} = \lambda \boldsymbol{w}$就是传统的做法，但不推荐，因为对于LDA问题，还有更快的做法

- 现在来看另一种求解 $\boldsymbol{w}$ 和 λ 的思路（Fisher提出来的）

3.5 Fisher的求解思路

- 再一次观察，由(16.1)式中的定义

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

- 那么(22)式中的

$$S_B\boldsymbol{w} = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T\boldsymbol{w} \quad (23)$$

3.5 Fisher的求解思路

- 由(23)，可以将(22) $S_W^{-1}S_B\boldsymbol{w} = \lambda\boldsymbol{w}$改写为

$$S_W^{-1}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T\boldsymbol{w} = \lambda\boldsymbol{w} \quad (24)$$

3.5 Fisher的求解思路

- (24)式中部分项可以“约掉”，得到

$$S_W^{-1}(\mu_1-\mu_2)(\mu_1-\mu_2)^T\boldsymbol{w}=\lambda\boldsymbol{w}\quad(24)$$

$$\Rightarrow\quad S_W^{-1}(\mu_1-\mu_2)=\boldsymbol{w}\quad(25)$$

- Question：为什么这里可以如此“约掉”？

3.5 Fisher的求解思路

- 现在来分析为什么可以如此“约掉”

- 看前面的(16.1)式$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$，不难发现，$S_B\boldsymbol{w}$，即$(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T\boldsymbol{w}$，与$\mu_1 - \mu_2$的方向总是一致的

- (24)式：$S_W^{-1}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T\boldsymbol{w} = \lambda\boldsymbol{w}$的左边可以写成$S_W^{-1}(\mu_1 - \mu_2)$

Notice：蓝颜色的部分与紫颜色的部分方向一致

3.5 Fisher的求解思路

- 再观察(24)式：$S_W^{-1}(\mu_1-\mu_2)(\mu_1-\mu_2)^T\boldsymbol{w}=\lambda\boldsymbol{w}$

- 看右边$\lambda\boldsymbol{w}$，它的方向与$\boldsymbol{w}$一致，因此可以写为$\boldsymbol{w}$

- 这样，(24)式就变为

$$S_W^{-1}(\mu_1-\mu_2)=\boldsymbol{w} \quad (25)$$

3.5 Fisher的求解思路

根据(25)式，可得

$$\boldsymbol{w} = S_W^{-1}(\mu_1 - \mu_2) \quad (26)$$

其中，

$$S_W = S_1 + S_2 = S_i$$

$$= \sum_{y \in \omega_1} (x - \mu_1)(x - \mu_1)^T + \sum_{y \in \omega_2} (x - \mu_2)(x - \mu_2)^T$$

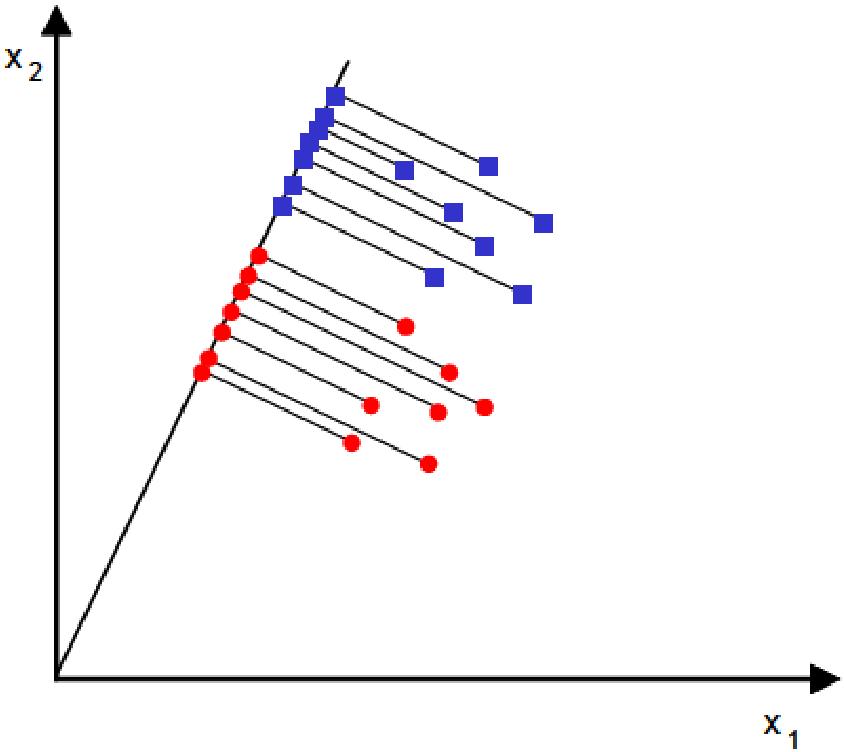
跟协方差矩阵相比，它仅仅没有除以样本数目，这样的矩阵S_i称为散列矩阵

3.5 Fisher的求解思路

- 综上所述，根据Fisher的思想，要求出$\boldsymbol{w}$，只需要求出原始样本的均值和方差就可以了

- 下面感性理解为什么Fisher可以这么做？

3.5 Fisher的求解思路



- 前面说过，我们要找的投影向量$\boldsymbol{w}$，相当于一个射线

- 对于这条射线，我们并不关心它的大小，我们只关心它的方向

4．多元LDA线性判别分析

4.1 多元LDA问题引入

- 前面描述的只是两个类情况下的LDA算法，在实际人脸识别中，如果有C个人的样本，那么类别的数量一定是C， C肯定不止两个

4.1 多元LDA问题引入

- 另一方面，之前进行二元LDA时，我们原有的数据是二维的

- 我们希望找到一个合适的向量$\boldsymbol{w}$进行投影，将原有的二维数据用一维的量（比如到圆点的距离）来表示而且区分开它们

- 而现在类别多了，降到一维可能不够用

4.1 多元LDA问题引入

- 多元LDA问题：数据有C个类别，需要K维向量（或者叫做基向量）来进行投影

- 这K维向量表示为$W=[\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_K]$，将样本点在这$K$维向量投影后结果表示为$[\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_K]$，满足

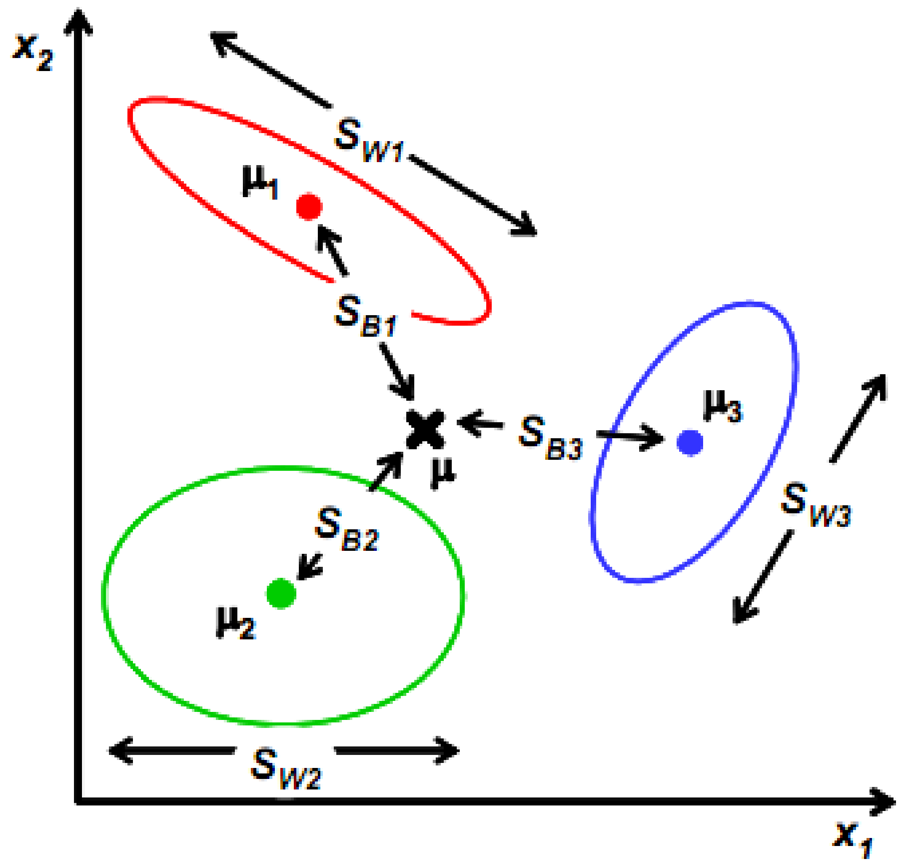
$$y_i=\boldsymbol{w}_i^T x \quad (27)$$

$$y=W^T x \quad (28)$$

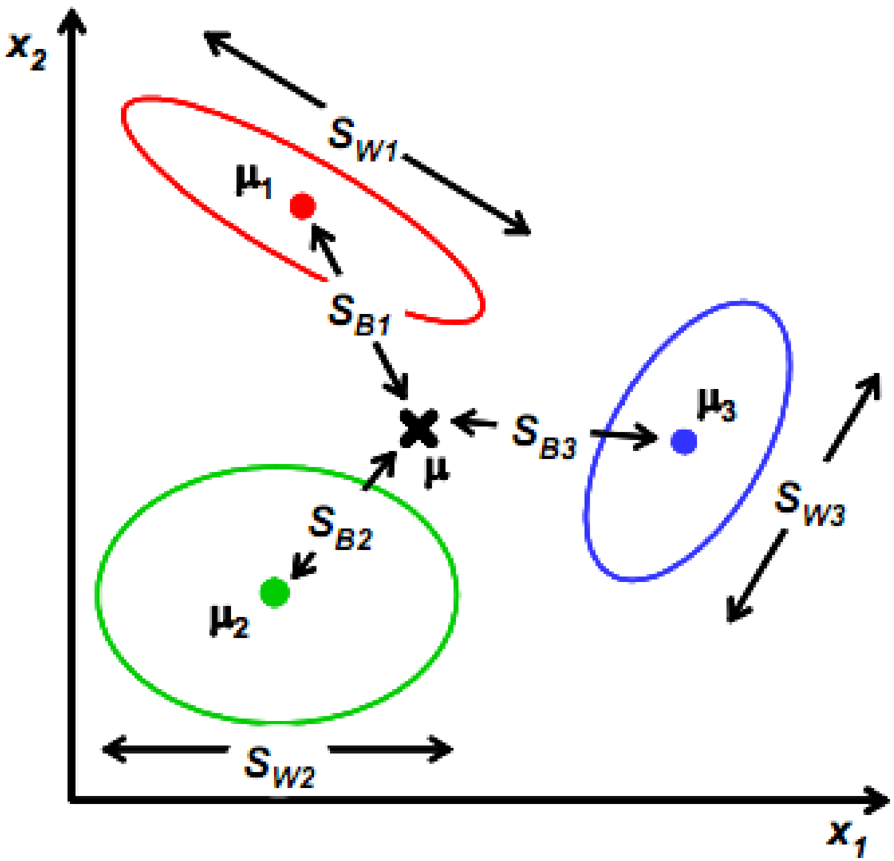
4.2 概念定义

- 为了像上节一样度量$J(\boldsymbol{w})$，我们打算仍然从类间散列度和类内散列度来考虑。

- 当样本是二维时，我们从几何意义上考虑：

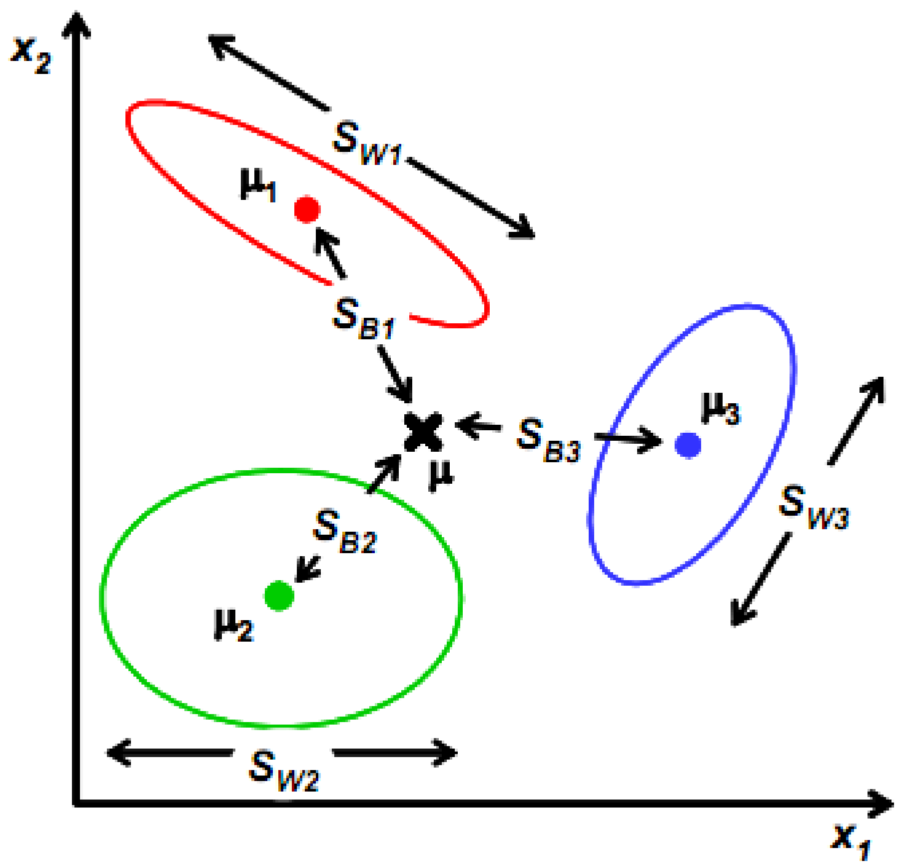


4.2 概念定义



- 其中μ_i和与S_W跟上节的意义一样，比如S_{W_1}是类别1里的样本点相对于该类中心点μ_1的散列程度。（此处的S_{W_1}应该就是前一节中的S_i）

4.2 概念定义



- S_{B_1}变成类别1中心点相对于所有样本中心点μ的协方差矩阵，即类别1相对于 μ 的散列程度。

4.2 概念定义

- 某类样本的类内散列矩阵S_{W_i}，它的维度是$n \times n$

$$S_{W_i} = \sum_{x \in \omega_i} (x - \mu_i)(x - \mu_i)^T \quad (29)$$

- 全体样本类内散列矩阵S_W

$$S_W = \sum_{i=1}^{C} S_{W_i} = \sum_{i=1}^{C} \sum_{x \in \omega_i} (x - \mu_i)(x - \mu_i)^T \ (30)$$

4.2 概念定义

- 所有样本的中心（N_i代表类别i内x的个数）

$$\mu = \frac{1}{N}\sum_{\forall x} x = \frac{1}{N}\sum_{\forall x} N_i \mu_i \quad (31)$$

- 原始样本类间散列矩阵<u>**加和**</u>S_B（N_i代表类别i内x的个数），注意多元情况下S_B的概念与二元情况下的不同

$$S_B = \sum_{i=1}^{C} N_i(\mu_i - \mu)(\mu_i - \mu)^T \quad (32)$$

4.2 概念定义

- 从(32)式可以看到，在多元情况下，S_B的计算发生了改变，对比二元情况下的(16)式

$$S_{B(二元)} = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \quad (16)$$

$$S_{B(多元)} = \sum_{i=1}^{C} N_i(\mu_i - \mu)(\mu_i - \mu)^T \quad (32)$$

4.2 概念定义

- 对比上述(16)式和(32)式，二元情况下S_B度量的是两个均值点的散列情况

- 多元情况下S_B度量的是每类均值点相对于样本中心的散列情况

4.2 概念定义

- 总结一下概念的变化

	二元LDA	多元LDA
μ_i	某类样本中心点	某类样本中心点
S_{W_i}	类内散列矩阵	类内散列矩阵
S_B	（S_B）类间散列矩阵	（S_{B_i}）类间散列矩阵和二元情况下有较大不同，它代表的是每个类别到μ距离的**加和**

4.2 概念定义

- 某一类样本投影后的均值

$$\tilde{\mu}_i = \frac{1}{N_i} \sum_{y \in \omega_i} y = \frac{1}{N_i} \sum_{x \in \omega_i} W^T x \quad (33)$$

- 所有样本投影后的均值

$$\tilde{\mu} = \frac{1}{N} \sum_{\forall y} y = \frac{1}{N} \sum_{\forall x} W^T x \quad (34)$$

4.2 概念定义

- 投影后的类内散列矩阵$\widetilde{S_W}$

$$\widetilde{S_W} = \sum_{i=1}^{C} \sum_{y \in \omega_i} (y - \widetilde{\mu_i})(y - \widetilde{\mu_i})^T$$

$$= \sum_{i=1}^{C} \sum_{x \in \omega_i} (W^T x - \widetilde{\mu_i})(W^T x - \widetilde{\mu_i})^T$$

(35)

4.2 概念定义

- 与前述(30)式比较

$$S_W = \sum_{i=1}^{C} \sum_{x \in \omega_i} (x - \mu_i)(x - \mu_i)^T \quad (30)$$

$$\widetilde{S_W} = \sum_{i=1}^{C} \sum_{x \in \omega_i} (W^T x - \widetilde{\mu_i})(W^T x - \widetilde{\mu_i})^T \quad (35)$$

4.2 概念定义

- 投影后的类间散列矩阵$\widetilde{S_B}$

$$\widetilde{S_B} = \sum_{i=1}^{C} N_i(\widetilde{\mu}_i - \widetilde{\mu})(\widetilde{\mu}_i - \widetilde{\mu})^T \qquad (36)$$

其中，N_i代表类别i内x的个数

$\widetilde{\mu}_i$代表某一类样本投影后的均值

$\widetilde{\mu}$代表所有样本投影后的均值

4.2 概念定义

- 与前述(32)式比较

$$S_B = \sum_{i=1}^{C} N_i(\mu_i - \mu)(\mu_i - \mu)^T \quad (32)$$

$$\widetilde{S_B} = \sum_{i=1}^{C} N_i(\tilde{\mu}_i - \tilde{\mu})(\tilde{\mu}_i - \tilde{\mu})^T \quad (36)$$

4.2 概念定义

- 综合各个投影向量$\boldsymbol{w}$上的$\widetilde{S_W}$（投影后的类内散列矩阵）和$\widetilde{S_B}$（投影后的类间散列矩阵），更新这两个参数，得到

$$\widetilde{S_W} = W^T S_W W \quad (37)$$

$$\widetilde{S_B} = W^T S_B W \quad (38)$$

其中的推导见下页

4.2 概念定义

蓝色部分参考关于(16)式的推导及其结论

- 现在推导(37)式

$$W^T S_W W = W^T \sum_{i=1}^{C} \sum_{x \in \omega_i} (x - \mu_i)(x - \mu_i)^T W$$

$$= \sum_{i=1}^{C} \sum_{x \in \omega_i} W^T (x - \mu_i)(x - \mu_i)^T W = \sum_{i=1}^{C} \sum_{x \in \omega_i} (W^T x - W^T \mu_i)^2$$

$$= \sum_{i=1}^{C} \sum_{x \in \omega_i} (W^T x - \widetilde{\mu_i})^2 = \sum_{i=1}^{C} \sum_{x \in \omega_i} (W^T x - \widetilde{\mu_i})(W^T x - \widetilde{\mu_i})^T = \widetilde{S_W}$$

4.2 概念定义

- (38)式推导同上，此处略

- 经过推导后，现在我们完全能够确认(37)和(38)式的合理性

4.3 求解标准

投影后的类
间散列矩阵

- 仿照二元LDA的求解标准(17)式，在多元情况下，根据前面推导过的(37)式和(38)式，定义

$$J(W)=\frac{|\widetilde{S_B}|}{|\widetilde{S_W}|}=\frac{|W^T S_B W|}{|W^T S_W W|} \quad (39)$$

投影后的类
内散列矩阵

- 但是，表达式中为什么会有|　|的符号？答案是：行列式记号，后面详细解释

4.3 求解标准

- 将(17)式粘贴过来对照着看

$$J(\boldsymbol{w}) = \frac{|\widetilde{\mu_1} - \widetilde{\mu_2}|^2}{\widetilde{s_1}^2 + \widetilde{s_2}^2} = \frac{\boldsymbol{w}^T S_B \boldsymbol{w}}{\boldsymbol{w}^T S_W \boldsymbol{w}} \quad (17)$$

二元LDA

$$J(W) = \frac{|\widetilde{S_B}|}{|\widetilde{S_W}|} = \frac{|W^T S_B W|}{|W^T S_W W|} \quad (39)$$

多元LDA

注意：(17)式中的$\boldsymbol{w}$是一个向量，(39)式中的W是一个矩阵，刚刚提到的|　|符号问题是否与这种差异相关？

4.3 求解标准

- 关于(17)式的分子，我们回顾(16)式，可以肯定$\boldsymbol{w}^T S_B \boldsymbol{w}$的结果是一个标量

$$|\widetilde{\mu_1} - \widetilde{\mu_2}|^2 = (\widetilde{\mu_1} - \widetilde{\mu_2})^2 = (\boldsymbol{w}^T \mu_1 - \boldsymbol{w}^T \mu_2)^2$$
$$= \boldsymbol{w}^T (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \boldsymbol{w} = \boldsymbol{w}^T S_B \boldsymbol{w} \quad (16)$$

- 关于(17)式的分母，我们回顾(15)式，可以肯定$\widetilde{s_1}^2 + \widetilde{s_2}^2$的计算结果是一个标量

$$\widetilde{s_1}^2 + \widetilde{s_2}^2 = \boldsymbol{w}^T S_1 \boldsymbol{w} + \boldsymbol{w}^T S_2 \boldsymbol{w} = \boldsymbol{w}^T (S_1 \boldsymbol{w} + S_2 \boldsymbol{w})$$
$$= \boldsymbol{w}^T [(S_1 + S_1) \boldsymbol{w}] = \boldsymbol{w}^T (S_1 + S_1) \boldsymbol{w} = \boldsymbol{w}^T S_W \boldsymbol{w} \quad (15)$$

4.3 求解标准

- 所以，(17)式的表达式实质上是两个标量相除，这也是二元LDA计算的特点

矩阵

- 可是多元LDA就不一样了，看(39)式

$$J(W)=\frac{|\widetilde{S_B}|}{|\widetilde{S_W}|}=\frac{|W^T S_B W|}{|W^T S_W W|} \quad (39)$$

分子和分母|　|符号里面的内容都是矩阵，而矩阵不能直接做除法，因此我们需要引入某种机制，这里引入的是行列式（是否还存在其他办法尚不了解）

4.3 求解标准

- 理由1：要将矩阵变成实数，需要取行列式

下面是引用的网页资料截图

然而，最后的J(w)的形式是

$$J(w)=\frac{|\widetilde{S_B}|}{|\widetilde{S_w}|}=\frac{|W^TS_BW|}{|W^TS_wW|} \quad (13)$$

由于得到的分子分母都是散列矩阵，要将矩阵变成实数，需要取行列式。又因为行列式的值实际上是矩阵特征值的积，一个特征值可以表示在该特征向量上的发散程度。因此我们使用行列式来计算。

4.3 求解标准

- 理由2：又因为行列式的值实际上是矩阵特征值的积，一个特征值可以表示在该特征向量上的发散程度。

下面是引用的网页截图

? 为什么矩阵的行列式等于他所有特征值的乘积

为什么矩阵的行列式等于他所有特征值的乘积

幽灵辉耀团3339　数学　2014-11-11

优质解答　　下载作业帮App，拍照秒答

因为矩阵可以化成对角元素都是其特征值的对角矩阵,而行列式的值不变,对角矩阵的行列式就是对角元素相乘

lwblXYbd　2014-11-11

4.3 求解标准

- 经过上述推导和分析，我们已经能够确认前述(39)式的合理性

$$J(W)=\frac{|\widetilde{S_B}|}{|\widetilde{S_W}|}=\frac{|W^T S_B W|}{|W^T S_W W|} \quad (39)$$

- 现在的问题是：求出W使$J(W)$能够取得最大值

4.4 求解

- W 由下式最大特征值对应的特征向量组成

$$S_B \boldsymbol{w}_i = \lambda_i S_W \boldsymbol{w}_i \quad (40)$$

注意：关于为什么可以这么做，权威书籍说其中的证明比较复杂（包括《模式分类》和《模式识别与机器学习》），均没有给出过程，我们只需知道这么用就够了

4.4 求解

- 对上述(40)式$S_B\boldsymbol{w}_i = \lambda_i S_W \boldsymbol{w}_i$求特征向量

- 首先，如果S_W是非奇异的，那么(40)式的求解实质上就是传统特征向量的求解，然而这么做其实并不方便，因为在对S_W求逆时要进行不必要的计算。相应地，我们可以将特征值作为下述(41)的根进行求解

4.4 求解

- 特征值可以看作是(41)式的根

$$|S_B - \lambda_i S_W| = 0 \quad (41)$$

- 得到了特征值，可以根据前述(40)式直接求出对应的特征向量

$$S_B \boldsymbol{w}_i = \lambda_i S_W \boldsymbol{w}_i \quad (40)$$

4.4 求解

- 回顾(32)式

$$S_B = \sum_{i=1}^{C} N_i(\mu_i - \mu)(\mu_i - \mu)^T \quad (32)$$

- 注意：由于$(\mu_i - \mu)$秩为1，因此S_B的秩至多为C（矩阵的秩小于等于各个相加矩阵的秩的和）。由于知道了前$C-1$个μ_i后，最后一个μ_C可以有前面的$C-1$个μ_i线性表示，因此S_B的秩至多为$C-1$。那么特征向量最多有$C-1$个

汇报结束，谢谢大家！