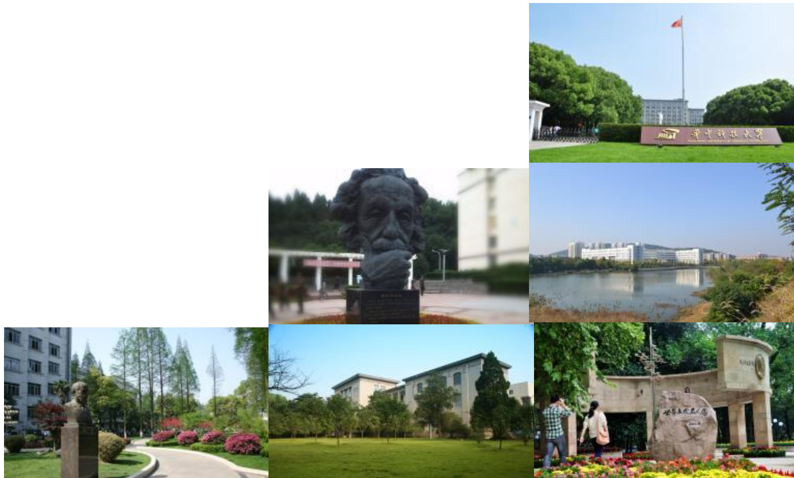
模式识别

特征提取



第六章　特征提取(Feature Extraction)

6.1　基于类别可分离性判据的特征提取

6.2　主成分分析

6.3　K-L变换方法

引言Introduction

特征提取：把D个特征通过适当的变换转换为d (D>d)个新特征，实现特征空间降维，也称特征变换或特征压缩。

（这里只讨论线性变换）

若$\{\vec{X}\}$是ω_i类的D 维样本集，寻找一个矩阵 W，通过变换 $\vec{Y}=W^T\vec{X}$，将$\vec{X}$ 压缩成 d 维向量$\vec{Y}$ (*d*< D)。

特征提取变换矩阵Transform Matrix

$\vec{Y}=(y_1,y_2,\cdots,y_d)'$ $\Rightarrow$ $\underbrace{\vec{Y}}_{d\times 1}=\underbrace{W^T}_{d\times D}\underbrace{\vec{X}}_{D\times 1}$ $\Leftarrow$ $\vec{X}=(x_1,x_2,\cdots,x_D)'$

二次特征 **原始特征**

(*d*< D)

注意：维数降低后，在新的d 维空间里各模式类之间的分布规律应至少保持不变或更优化，即压缩维数同时保留类别间的鉴别信息，突出可分性。

6. 1 基于类别可分离性判据(Two-Class Separability Measures) 的特征提取

准则函数：$J(W)$（变换后的可分离性判据）

$$S_W^* = W^T S_W W$$

$$S_B^* = W^T S_B W$$

$$J_1(W) = \mathrm{tr}\left[W^T(S_w + S_b)W\right]$$

$$J_2(W) = \mathrm{tr}\left[\left(W^T S_w W\right)^{-1}\left(W^T S_b W\right)\right]$$

$$J_3(W) = \ln\frac{\left|W^T S_b W\right|}{\left|W^T S_w W\right|}$$

$$J_4(W) = \frac{\mathrm{tr}(W^T S_b W)}{\mathrm{tr}(W^T S_w W)}$$

$$J_5(W) = \frac{\left|W^T \Sigma W\right|}{\left|W^T S_w W\right|} \qquad \Sigma = S_w + S_b$$

$$J_1 = tr\left(S_W + S_B\right)$$

$$J_2 = \mathrm{tr}\left(S_w^{-1} S_b\right)$$

$$J_3 = \ln\frac{\left|S_b\right|}{\left|S_w\right|}$$

$$J_4 = \frac{\mathrm{tr}S_b}{\mathrm{tr}S_w}$$

$$J_5 = \frac{\left|S_b - S_w\right|}{\left|S_w\right|}$$

目标：求 W^*，使 $J(y) = \max\limits_{\{w\}} J(W^T x)$

矩阵求导(Taking the derivative of the matrix)

a 是实数，β，X 是向量，A, B, C 是与 X 无关的矩阵

$$\frac{\partial \beta^T X}{X} = \beta$$

$$\frac{\partial X^T X}{\partial X} = X$$

$$\frac{\partial X^T A X}{\partial X} = (A + A^T)X$$

Andrew Ng 推荐的使用矩阵的迹的相关公式：

$\mathrm{tr}(a) = a$

$\mathrm{tr}(AB) = \mathrm{tr}(BA)$

$\mathrm{tr}(ABC) = \mathrm{tr}(CAB) = \mathrm{tr}(BCA)$

$$\frac{\partial tr(AB)}{\partial A} = B^T$$

$\mathrm{tr}(A) = \mathrm{tr}(A^T)$

$$\frac{\partial tr(ABA^T C)}{\partial A} = CAB + C^T AB^T$$

$$\boxed{J_1(W) = \mathrm{tr}\left(W^T (S_w + S_b) W\right)}$$

$$\frac{\partial J_1(W)}{\partial W} = \frac{\partial \mathrm{tr}\left(W^T (S_w + S_b) W\right)}{\partial W}$$

$$= \frac{\partial \mathrm{tr}\left(W W^T (S_w + S_b)\right)}{\partial W}$$

$$= (S_w + S_b)W + (S_w + S_b)^T W$$

$$= 2(S_w + S_b)W$$

矩阵求导(Taking the derivative of the matrix)

a 是实数，β，X 是向量，A, B, C 是与 X 无关的矩阵

$$\frac{\partial \beta^T X}{X} = \beta$$

$$\frac{\partial X^T X}{\partial X} = X$$

$$\frac{\partial X^T A X}{\partial X} = (A + A^T)X^*$$

$$S_W^* = W^T S_W W$$

Andrew Ng 推荐的使用矩阵的迹的相关公式：

$$S_B^* = W^T S_B W$$

$\mathrm{tr}(a) = a$

$\mathrm{tr}(AB) = \mathrm{tr}(BA)$

$\mathrm{tr}(ABC) = \mathrm{tr}(CAB) = \mathrm{tr}(BCA)$

$$\frac{\partial tr(AB)}{\partial A} = B^T$$

$\mathrm{tr}(A) = \mathrm{tr}(A^T)$

$$\frac{\partial tr(ABA^T C)}{\partial A} = CAB + C^T AB^T$$

$$tr\left[\Lambda(W^T S_w W - I)\right] = tr\left[\Lambda W^T S_w W - \Lambda\right]$$

$$= tr\left[\Lambda W^T S_w W\right] - tr[\Lambda]$$

$$\frac{\partial tr\left[\Lambda(W^T S_w W - I)\right]}{\partial W} = \frac{\partial tr\left(\Lambda W^T S_w W\right)}{\partial W}$$

$$tr\left(\Lambda W^T S_w W\right) = tr\left(W^T S_w{}^T W \Lambda^T\right)$$

$$tr\left(W^T S_w{}^T W \Lambda^T\right) = tr\left(W \Lambda^T W^T S_w{}^T\right)$$

$$\frac{\partial tr\left(W \Lambda^T W^T S_w{}^T\right)}{\partial W} = S_w{}^T W \Lambda^T + S_w W \Lambda$$

$$= 2 S_w W \Lambda$$

矩阵迹(Matrix Trace)形式的判据

$J_2 = Tr\left[S_W^{-1}S_B\right]$

以 J_2 为例

$$J_2^*(W) = Tr\left[\left(S_W^*\right)^{-1} S_B^*\right] = Tr\left[\left(W^T S_W W\right)^{-1} W^T S_B W\right]$$

$P^{-1}AP = B$, A ~ B,tr(A)=tr(B)

$W_e^{-1} = W_e^{T}$

由线性代数可知，对矩阵作相似变换(similar transform)其迹不变，一个方阵的迹等于它的所有特征值(Eigenvalues)之和。设 W_e 为正交阵(Orthogonal matrix)，用 W_e 对对称阵(Symmetric matrix) $S_W^{-1}S_B$ 作相似变换使其成为对角阵(Diagonal matrix)

相似变换实对称阵可对角化

$$W_e^{-1}S_W^{-1}S_B W_e = W_e^{T}S_W^{-1}S_B W_e = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_D \end{pmatrix} = diag(\lambda_1, \lambda_2, \cdots, \lambda_D)$$

其中，λ_i (i=1,2,…D)为 $S_W^{-1}S_B$ 的特征值，W_e 的列矢量(Column vector) 为 $\vec{W}_i$ $S_W^{-1}S_B$ 相应于λ_i的特征矢量(Eigenvector)。由上式可得。

$$J_2 = Tr[S_W^{-1}S_B] = Tr[W_e^T S_W^{-1}S_B W_e] = \sum_{i=1}^{D}\lambda_i$$

矩阵迹(Matrix Trace)形式的判据

转换步骤：使W_e的列矢量的排列作适当调整，使$S_W^{-1}S_B$的特征值$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_D$。由此可得，在d给定后，取前d个较大的特征值所对应的特征矢量w_i(i=1, 2, …, d)构造特征提取矩阵(transform matrix)W，即：

$$W=(\vec{w}_1 \quad \vec{w}_2 \quad \cdots \quad \vec{w}_d)$$

$D \times d$阵

$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d \geq \ldots \geq \lambda_D$

作变换 $\vec{y}=W^{T}\vec{x}$ **这时对于给定的*d*所得到的**

$$J_2^*(W)=\sum_{i=1}^{d}\lambda_i$$

达最大值。

行列式(Determinant)形式的判据

$$J_5 = \frac{|\Sigma|}{|S_W|}$$

所有特征值大于0

行列式不为0，可逆

以J_5为例，由于S_W是对称正定矩阵(symmetric, and positive-definite matrix)，设有非奇异阵(Nonsingular matrix)A，使

$$A^T S_W A = I$$

正定矩阵和单位矩阵(Identity matrix)合同(Congruent transformation)

设有标准正交矩阵(Orthonormal matrix)V，使 $V^{-1} A^T \Sigma A V = \tilde{\Lambda}$

相似变换实对称阵可对角化(Diagonalization)

这里 $\tilde{\Lambda}$ **为对角阵(Diagonal matrix)。另外有** $V^{-1} A^T S_W A V = V^{-1} I V = I$

令$U=AV$，因此存在非奇异矩阵U，使

$$U^T \Sigma U = \begin{pmatrix} \tilde{\lambda}_1 & & & 0 \\ & \tilde{\lambda}_2 & & \\ & & \ddots & \\ 0 & & & \tilde{\lambda}_D \end{pmatrix} = \tilde{\Lambda} \qquad U^T S_W U = I$$

行列式(Determinant)形式的判据

$$U^T \sum U = \begin{pmatrix} \tilde{\lambda}_1 & & & 0 \\ & \tilde{\lambda}_2 & & \\ & & \ddots & \\ 0 & & & \tilde{\lambda}_D \end{pmatrix} = \tilde{\Lambda}$$

$$U^{\mathrm{T}} S_W U = I$$

上面右式两边同时取逆(Inverse)，有 $U^{-1} S_w^{\ -1} (U^{\mathrm{T}})^{-1} = I$

再与左式相乘，并左乘*U*(Pre-multiply it by U)，有:

$$S_W^{\ -1} \sum U = U \tilde{\Lambda}$$

***这里U*及$\tilde{\Lambda}$ 分别是特征矢量组成的矩阵(特征矢量矩阵Eigenvector matrix)及特征值对角阵(Diagonal matrix of eigenvalues)。**

行列式(Determinant)形式的判据

因为 $$S_W^{-1}\Sigma = S_W^{-1}(S_W + S_B) = I + S_W^{-1}S_B$$

所以 $$(I + S_W^{-1}S_B)U = U\tilde{\Lambda}$$ $$S_W^{-1}\Sigma U = U\tilde{\Lambda}$$

$$S_W^{-1}S_B U = U(\tilde{\Lambda} - I)$$

设$S_W^{-1}S_B$ 的特征值对角矩阵 $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_D)$

则 $I + \Lambda = \tilde{\Lambda}$

$$U^{-1}S_W^{-1}S_B U = U^{-1}U(\tilde{\Lambda} - I) = \begin{pmatrix} \tilde{\lambda}_1 - 1 & & & 0 \\ & \tilde{\lambda}_2 - 1 & & \\ & & \ddots & \\ 0 & & & \tilde{\lambda}_D - 1 \end{pmatrix} = \tilde{\Lambda} - I = \Lambda$$ $$U = W_e$$

于是

$$J_5 = \frac{|\Sigma|}{|S_W|} = |S_W^{-1}||\Sigma| = |S_W^{-1}\Sigma| = |U^{-1}S_W^{-1}\Sigma U| = \prod_{i=1}^{D}\tilde{\lambda}_i = \prod_{i=1}^{D}(1+\lambda_i)$$

相似变换行列式不变

行列式(Determinant)形式的判据

$$J_5=\frac{|\Sigma|}{|S_W|}=\left|S_W^{-1}\right||\Sigma|=\left|S_W^{-1}\Sigma\right|=\left|U^{-1}S_W^{-1}\Sigma U\right|=\prod_{i=1}^{D}\tilde{\lambda}_i=\prod_{i=1}^{D}(1+\lambda_i)$$

此处的*U* 就是前述的*W*$_e$。

转换步骤：设*U* 的各列已作适当调整，使S$_W^{-1}$S$_B$ 的特征值$\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_D$，对于给定的*d*，取前*d*个较大的特征值对应的特征矢量构造变换矩阵可使J_5取最大值，此时

$$J_5^*(W)=\prod_{i=1}^{d}(1+\lambda_i)$$

结论：设矩阵 $S_w^{-1}S_b$的特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_D$，且 $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_D$

选前 d个特征值对应的特征向量 u_1，u_2，…，u_d 组成矩阵

$$W=\left[u_1\ u_2\ \cdots\ u_d\right]$$

就是上述准则下的最佳变换阵。

6.2 主成分分析 Principal Component Analysis

- **PCA的思想**

将 n 维特征映射到k维上（$k<n$），这 k 维是全新的正交特征，称为主元，这是一组按重要性从大到小排列的新特征，它们是原有特征的线性组合，并且相互之间是不相关的。

- **PCA 方法**

Example:

Two Dimensional Data—One Dimensional Data

Rows represent samples. Columns represent features. There are ten samples, and each sample has two features.

Data =

x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9

6.2 主成分分析 Principal Component Analysis

First Step： 均值归零（分别求各特征平均值，然后对于所有的samples的特征，减去其对应的均值）

这里x的均值是1.81，y的均值是1.91，那么第一个*sample*减去均值后即为（0.69,0.49），得到

Data =

x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9

DataAdjust =

x	y
.69	.49
-1.31	-1.21
.39	.99
.09	.29
1.29	1.09
.49	.79
.19	-.31
-.81	-.81
-.31	-.31
-.71	-1.01

6.2 主成分分析 Principal Component Analysis

Second Step：<u>求特征协方差矩阵</u>

如果数据是3维，那么协方差矩阵是

$$C = \begin{pmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{pmatrix}$$

$$\mathrm{cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

这里只有 *x* 和 *y*，求解得

$$cov = \begin{pmatrix} .616555556 & .615444444 \\ .615444444 & .716555556 \end{pmatrix}$$

6.2 主成分分析 Principal Component Analysis

Third Step：<u>求协方差的特征值和特征向量</u>

$|\lambda I-C|=0$

$$eigenvalues = \begin{pmatrix} .0490833989 \\ 1.28402771 \end{pmatrix}$$

$\lambda_i P_i = C P_i$

$$eigenvectors = \begin{pmatrix} -.735178656 & -.677873399 \\ .677873399 & -.735178656 \end{pmatrix}$$

上面是两个特征值，下面是对应的特征向量，特征值**0.0490833989**对应特征向量为 $(-0.735178656, 0.677873399)^T$ 。

Fouth Step：<u>将特征值按照从大到小的顺序排序，选择其中最大的k个，然后将其对应的k个特征向量分别作为列向量组成特征向量矩阵。</u>

这里特征值只有两个，我们选择其中最大的那个，这里是**1.28402771**，对应的特征向量是 $(-0.677873399, -0.735178656)^T$

6.2 主成分分析 Principal Component Analysis

Fifth Step：<u>将样本点投影到选取的特征向量上。假设样例数为m，特征数为n，减去均值后的样本矩阵为DataAdjust(m*n)，协方差矩阵是n*n，选取的k个特征向量组成的矩阵为EigenVectors(n*k)。那么投影后的数据FinalData为</u>

$$\boxed{\mathrm{FinalData}(m*k) = \mathrm{DataAdjust}(m*n) \times \mathrm{EigenVectors}(n*k)}$$

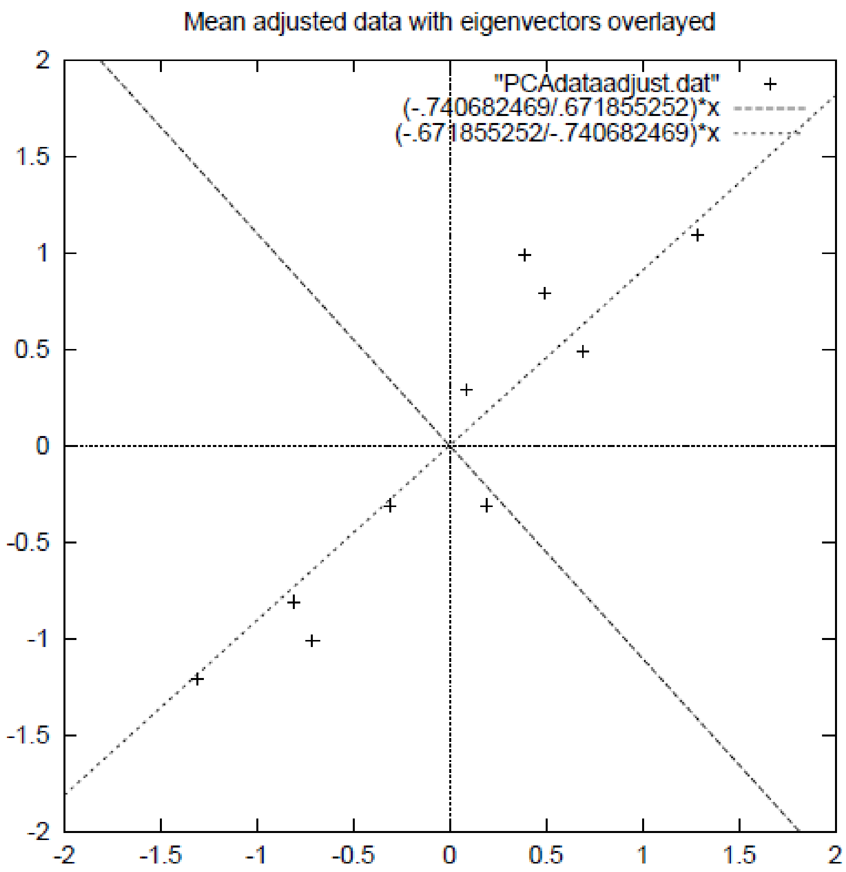
FinalData(10*1) = DataAdjust(10*2矩阵)×特征向量$(-0.677873399, -0.735178656)^T$

Transformed Data (Single eigenvector)

x
-.827970186
1.77758033
-.992197494
-.274210416
-1.67580142
-.912949103
.0991094375
1.14457216
.438046137
1.22382056

- 这样，就将原始样例的n维特征变成了k维，这k维就是原始特征在k维上的投影。
- 该特征基本上代表了原来的两个特征。

6.2 主成分分析 Principal Component Analysis



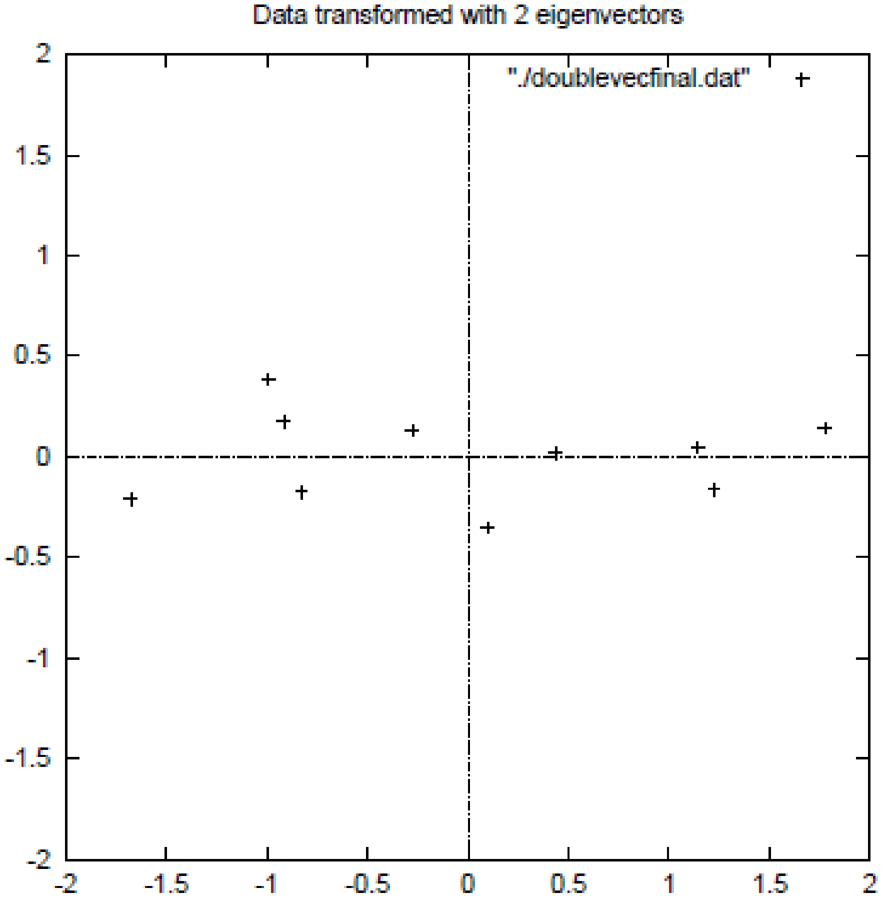
正号表示预处理后的样本点，斜着的两条线就分别是正交的特征向量（由于协方差矩阵是对称的，因此其特征向量正交），最后一步的矩阵乘法就是将原始样本点分别往特征向量对应的轴上做投影。

6.2 主成分分析 Principal Component Analysis

如果取的k=2，那么结果是

Transformed Data=

x	y
-.827970186	-.175115307
1.77758033	.142857227
-.992197494	.384374989
-.274210416	.130417207
-1.67580142	-.209498461
-.912949103	.175282444
.0991094375	-.349824698
1.14457216	.0464172582
.438046137	.0177646297
1.22382056	-.162675287



- 水平轴基本上可以代表全部样本点。整个过程看起来就像将坐标系做了旋转。
- 上面的如果**k=1**，那么只会留下这里的水平轴，轴上是所有点在该轴的投影。

第一步减均值之后，其实应该还有一步对特征做方差归一化。

比如一个特征是汽车速度（0到100），一个是汽车的座位数（2到6），显然第二个的方差比第一个小，因此，如果样本特征中存在这种情况，那么在第一步之后，求每个特征的标准差，然后对每个样例在该特征下的数据除以标准差。

归纳一下，在求协方差之前的步骤是：

1. Let $\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$.
2. Replace each $x^{(i)}$ with $x^{(i)} - \mu$.
3. Let $\sigma_j^2 = \frac{1}{m}\sum_i (x_j^{(i)})^2$
4. Replace each $x_j^{(i)}$ with $x_j^{(i)}/\sigma_j$.

其中$x^{(i)}$是样例，共m个，每个样例n个特征，也就是说$x^{(i)}$是n维向量。$x_j^{(i)}$是第i个样例的第j个特征。μ是样例均值。σ_j是第j个特征的标准差。

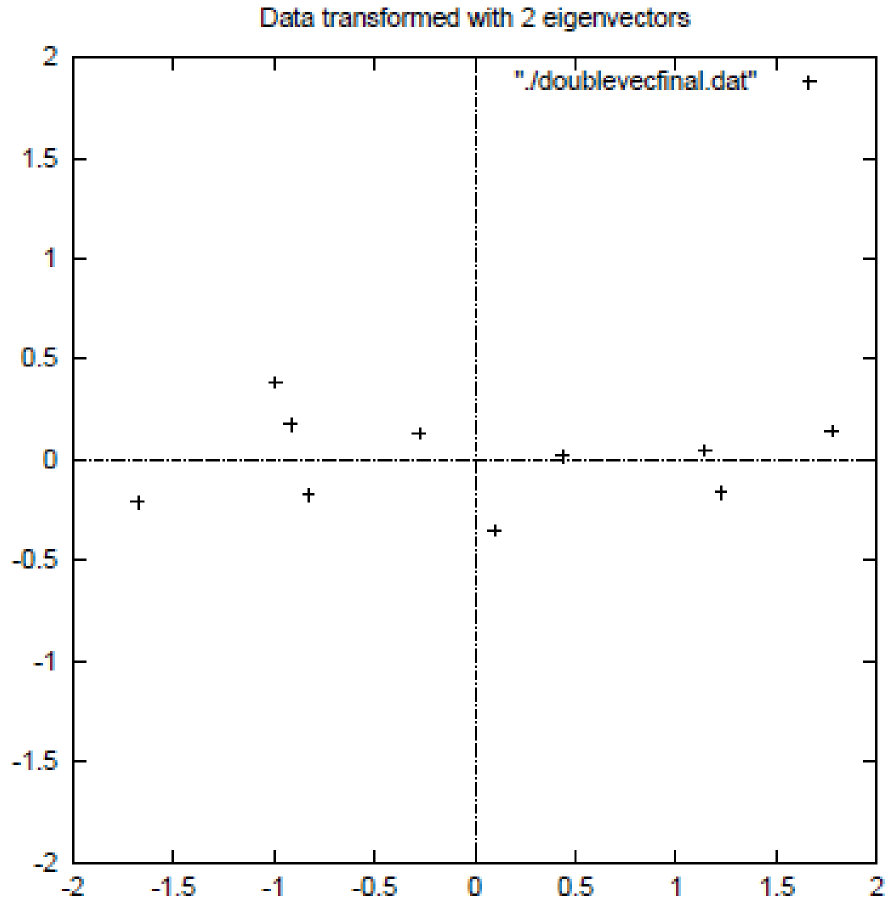
6.2 主成分分析 Principal Component Analysis

- 为什么协方差矩阵特征向量就是k维理想特征？

在信号处理中认为信号具有较大的方差，噪声有较小的方差。

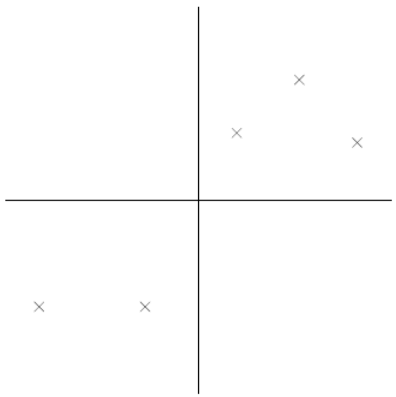
如图，样本在横轴上的投影方差较大，在纵轴上的投影方差较小，那么认为纵轴上的投影是由噪声引起的。

因此我们认为，<u>最好的 *k* 维特征是将 *n* 维样本点转换为 *k* 维后，每一维上的样本方差都很大。</u>



6.2 主成分分析 Principal Component Analysis

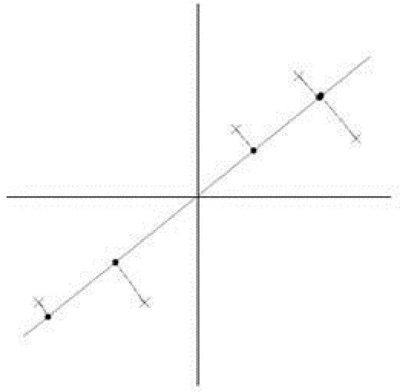
比如下图有5个样本点：（已经做过预处理，均值为0，特征方差归一）

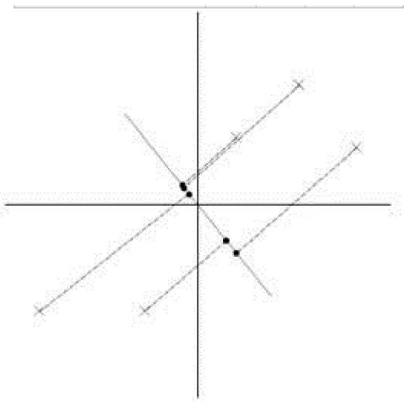


假设我们选择两条不同的直线做投影，那么左右两条中哪个好呢？

根据我们之前的方差最大化理论，左边的好，因为投影后的样本点之间方差最大。

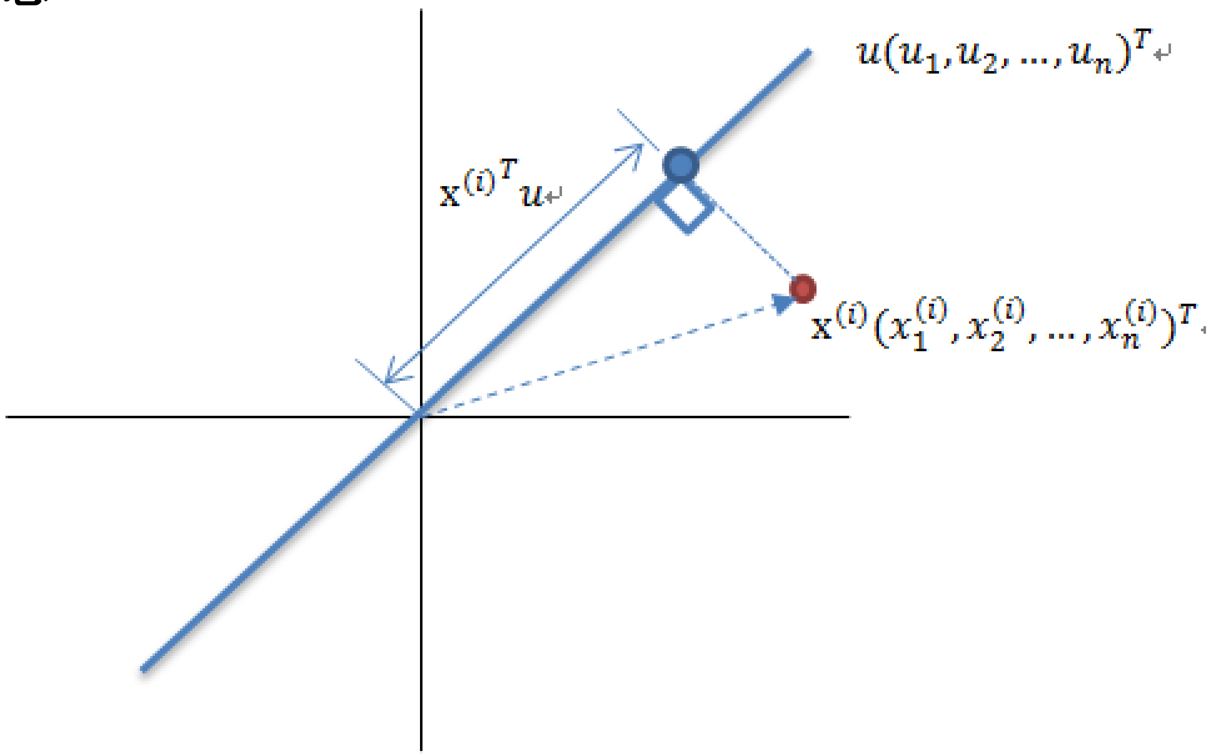
如果将样本投影到某一维上，这里用一条过原点的直线表示（前处理的过程实质是将原点移到样本点的中心点）。





6.2 主成分分析 Principal Component Analysis

投影的概念



红色点表示样例$\mathrm{x}^{(i)}$，蓝色点表示$\mathrm{x}^{(i)}$在u上的投影，u是直线的斜率也是直线的方向向量，而且是单位向量。蓝色点是$\mathrm{x}^{(i)}$在u上的投影点，离原点的距离是$< \mathrm{x}^{(i)}, u >$（即$\mathrm{x}^{(i)^T} u$或者$\mathrm{u}^T \mathrm{x}^{(i)}$）由于这些样本点（样例）的每一维特征均值都为0，因此投影到u上的样本点（只有一个到原点的距离值）的均值仍然是0。

6.2 主成分分析 Principal Component Analysis

我们要求的是**最佳的u，使得投影后的样本点方差最大**。

由于投影后均值为0，因此方差为：

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^{m}({x^{(i)}}^T u)^2 &= \frac{1}{m}\sum_{i=1}^{m} u^T x^{(i)} {x^{(i)}}^T u \\ &= u^T\left(\frac{1}{m}\sum_{i=1}^{m} x^{(i)} {x^{(i)}}^T\right)u.\end{aligned}$$

中间部分就是样本特征的协方差矩阵，**xi**的均值为0，一般协方差矩阵都除以m-1，这里用m。

用λ来表示$\frac{1}{m}\sum_{i=1}^{m}({x^{(i)}}^T u)^2$，$\Sigma$表示$\frac{1}{m}\sum_{i=1}^{m} x^{(i)} {x^{(i)}}^T$，那么上式写作

$\lambda = u^T \Sigma u$

由于u是单位向量，即$u^T u = 1$，上式两边都左乘u得，$u\lambda = \lambda u = u u^T \Sigma u = \Sigma u$

即$\Sigma u = \lambda u$

6.2 主成分分析 Principal Component Analysis

$$\Sigma u=\lambda u$$

λ就是∑的特征值，u是特征向量。最佳的投影直线是特征值λ最大时对应的特征向量，其次是λ第二大对应的特征向量，依次类推。

因此，我们只需要对协方差矩阵进行特征值分解，得到的前k大特征值对应的特征向量就是最佳的k维新特征，而且这k维新特征是正交的。得到前k个u以后，样本x_i通过以下变换可以得到新的样本。

$$y^{(i)}=\begin{bmatrix} u_1^T x^{(i)} \\ u_2^T x^{(i)} \\ \vdots \\ u_k^T x^{(i)} \end{bmatrix} \in \mathbb{R}^k.$$

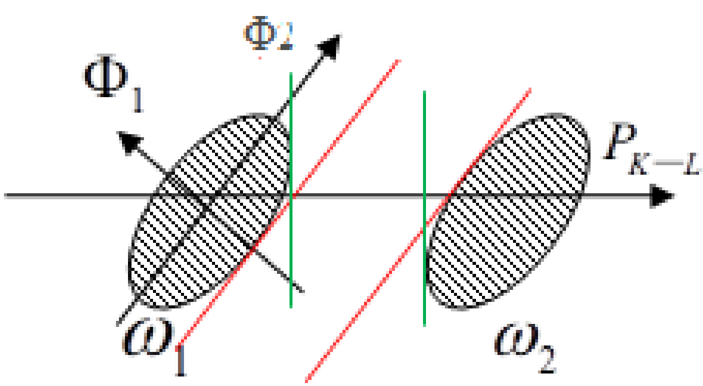
通过选取最大的**k**个**u**，使得方差较小的特征（如噪声）被丢弃。

其中的第j维就是$x^{(i)}$在u_j上的投影。

6.3 K-L变换方法Karhunen-Loeve transform

去相关性好

K-L变换：有多个变种，其最基本的形式原理上与主成分分析相同，同样是均方误差（MSE，MeanSquare Error）意义下的最佳变换。但K-L变换能够考虑到不同的分类信息。



$$P(\omega_1)=P(\omega_2)=0.5$$

- 如果降维对Φ_2轴投影，两类无法区分
- 如果对Φ_1方向投影，得到2类之间的距离即为2条红线之间的距离，但是这并不是相隔最远的投影方向。
- 将椭圆投影到P_{K-L}方向，得到2类之间的距离为2条绿线之间的距离。这个方向就是用自相关矩阵的统计平均得到的特征向量。

6.3 K-L变换方法Karhunen-Loeve transform

去相关性好

1. 最优描述的K-L变换（沿类间距离大的方向降维）

设共有C个类别，各类出现的先验概率为 $P(\omega_i) \quad i=1,\cdots C$

以x_i表示来自第i类的向量。则第i类集群的**自相关矩阵**为：$R_i = E\left(x_i x_i^T\right)$

混合分布的**自相关矩阵**R是：$R = \sum_{i=1}^{C} P(\omega_i) R_i = \sum_{i=1}^{C} P(\omega_i) E\left(x_i x_i^T\right)$

然后求出R的特征值和特征向量：$\Lambda = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_D \end{pmatrix} \quad \Phi = \left(\Phi_1, \Phi_2, \cdots \Phi_D\right)$

将特征值**降序**排列,为了降到d维，取前d个特征向量，构成变换矩阵A

$$A = \begin{pmatrix} \Phi_1^T \\ \vdots \\ \Phi_d^T \end{pmatrix}_{d\times D} \Rightarrow y = A_{d\times D} x_{D\times 1}$$

6.3 K-L变换方法Karhunen-Loeve transform

1. 最优描述的K-L变换（沿类间距离大的方向降维）

为什么K-L变换是均方误差（MSE，MeanSquare Error）意义下的最佳变换？

$$y^{(D)} = \Phi^T x \qquad x = \Phi \cdot y^{(D)} = \sum_{j=1}^{D} y_j^{(D)} \Phi_j$$

其中，$y_j^{(D)}$表示D维向量y的第j个分量,Φ_j表示第j个特征分量。

引起的均方误差为

$$e^2(m) = E\left\{\|\Delta x\|^2\right\} = E\left\{[\Delta x]^T[\Delta x]\right\} = E\left\{\left[\sum_{j=d+1}^{D} y_j^{(D)}\Phi_j\right]^T\left[\sum_{k=d+1}^{D} y_k^{(D)}\Phi_k\right]\right\}$$

$$= \sum_{j=d+1}^{n} E\left\{\left[y_j^{(D)}\right]^2\right\} = \sum_{j=d+1}^{D} E\left\{\left[\Phi_j^T x\right]^2\right\} = \sum_{j=d+1}^{D} \Phi_j^T R \Phi_j = \sum_{j=d+1}^{D} \lambda_j \qquad \Delta x = x - \hat{x} = \sum_{j=d+1}^{D} y_j^{(D)} \Phi_j$$

从d+1开始的特征值都是最小的几个，所以均方误差得到最小。

以上方法称为最优描述的K-L变换。沿类间距离大的方向降维，从而均方误差最佳。**最优描述的K-L变换扔掉了最不显著的特征，然而，显著的特征其实并不一定对分类有帮助。还是要找出对分类作用大的特征。**

2. 运用DKLT消除两类问题的特征相关性方法

（1） $y^{(1)}=U^{\mathrm{T}}x^{(1)}$ $\qquad x^{(1)}\in\omega_1$ $\qquad U$是S_{w1}（第一类的协方差）的特征矢量矩阵

$$\underline{U^T S_{W1} U=\Lambda_1=diag\left(\lambda_1^{(1)},\lambda_2^{(1)},\cdots\lambda_D^{(1)}\right) \qquad \leftarrow y^{(1)}\text{的各分量不相关}}$$

证明：

$$R_Y=E\left[YY^{\mathrm{T}}\right]=E\left[(U^TX)(U^TX)^T\right]=U^TR_XU=\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m\end{pmatrix}=\Lambda$$

$$\boxed{R_XU=\lambda U} \qquad R_XU=R_X\begin{bmatrix}u_1 & u_2 & \cdots & u_D\end{bmatrix}=\begin{bmatrix}\lambda_1u_1 & \lambda_2u_2 & \cdots & \lambda_Du_D\end{bmatrix}$$

$$U^TR_XU=\begin{bmatrix}u_1^T\\ u_2^T\\ \vdots\\ u_D^T\end{bmatrix}\begin{bmatrix}\lambda_1u_1 & \lambda_2u_2 & \cdots & \lambda_Du_D\end{bmatrix}=\begin{pmatrix}\lambda_1u_1^Tu_1 & & & 0\\ & \lambda_2u_2^Tu_2 & & \\ & & \ddots & \\ 0 & & & \lambda_Du_D^Tu_D\end{pmatrix}=\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D\end{pmatrix}$$

6.3 K-L变换方法Karhunen-Loeve transform

$$R_Y = E\left[YY^{\mathrm{T}}\right] = E\left[(U^T X)(U^T X)^T\right] = U^T R_X U = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{pmatrix} = \Lambda_{Rx}$$

$$C_Y = E\left[(Y - \overline{Y})(Y - \overline{Y})^T\right] = U^T C_X U = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{pmatrix} = \Lambda_{Cx}$$

6.3 K-L变换方法Karhunen-Loeve transform

2. 运用DKLT消除两类问题的特征相关性方法

称为白化变换

（2）再对$y^{(1)}$作变换： $\tilde{y}^{(1)}=\Lambda_1^{-1/2}y^{(1)}=\Lambda_1^{-1/2}U^{T}x^{(1)}\overset{def}{=}B^{T}x^{(1)}$

$$B=U\left(\Lambda_1^{-1/2}\right)^{T}=U\Lambda_1^{-1/2}$$ $\boxed{\text{白化变换矩阵}}$

求新变换后的协方差：

$$E\left[\left(\tilde{y}^{(1)}-\overline{\tilde{y}^{(1)}}\right)\left(\tilde{y}^{(1)}-\overline{\tilde{y}^{(1)}}\right)^{T}\right]=E\left[\left(\Lambda_1^{-1/2}y^{(1)}-\overline{\Lambda_1^{-1/2}y^{(1)}}\right)\left(\Lambda_1^{-1/2}y^{(1)}-\overline{\Lambda_1^{-1/2}y^{(1)}}\right)^{T}\right]$$

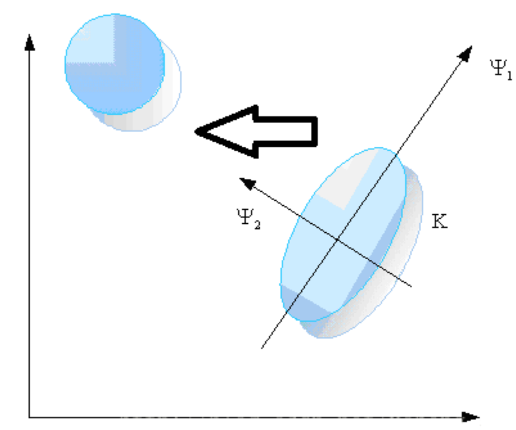
$$=\Lambda_1^{-1/2}E\left[\left(y^{(1)}-\overline{y^{(1)}}\right)\left(y^{(1)}-\overline{y^{(1)}}\right)^{T}\right]\left(\Lambda_1^{-1/2}\right)^{T}=\Lambda_1^{-1/2}\left(\Lambda_1\right)\left(\Lambda_1^{-1/2}\right)^{T}=I$$

即：$B^{T}S_{W1}B=\Lambda_1^{-1/2}U^{T}S_{W1}U\Lambda_1^{-1/2}=\Lambda_1^{-1/2}\Lambda_1\Lambda_1^{-1/2}=I$

这样$\tilde{y}^{(1)}$的协方差阵为单位阵，即各分量也不相关。

- 用U消除原来分量的相关性
- 用$\Lambda^{-1/2}$进行归一化

这一步是坐标尺度变换，相当于把椭圆整形成圆



6.3 K-L变换方法Karhunen-Loeve transform

2. 运用DKLT消除两类问题的特征相关性方法

（3）又再转换 $\tilde{S}_{W_2} = B^{\mathrm{T}} S_{W_2} B$

设V是$\tilde{S}_{W_2}$的特征矢量矩阵，令 $W^{\mathrm{T}} = V^{\mathrm{T}} B^{\mathrm{T}} = V^{\mathrm{T}} \Lambda_1^{-1/2} U^{\mathrm{T}}$

W作为最终的变换矩阵，变量x最终变换为向量z：

$$z = W^T x = V^{\mathrm{T}} \Lambda_1^{-1/2} U^{\mathrm{T}} x \qquad x \in \omega_1 \text{ 或 } x \in \omega_2$$

第2类样本协方差变换后$\tilde{S}_{W_2}$的特征矢量阵

$y^{(1)}$的协方差矩阵

第1类协方差$S_{\omega 1}$的特征矢量阵

上式z的**各分量是不相关**的。这样经变换后的各分量互不相关，于是可根据某种准则选择z分量以降低维数。

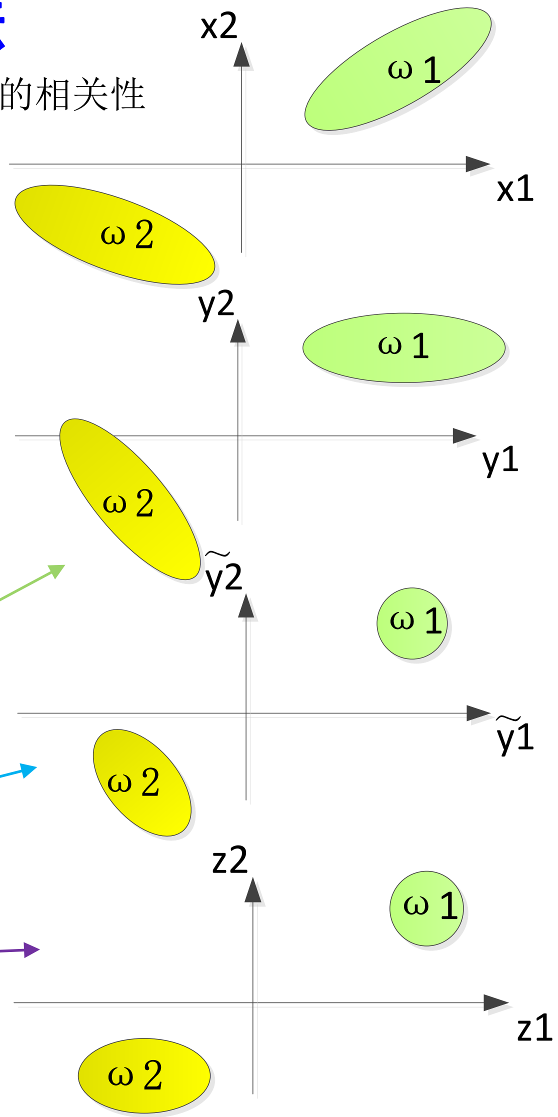
$$z^T z = x^T U \left(\Lambda_1^{-1/2}\right)^T V V^T \Lambda_1^{-1/2} U^T x = x^T U \Lambda_1^{-1/2} I \Lambda_1^{-1/2} U^T x$$

$$= x^T U \Lambda_1^{-1} U^T x = x \begin{pmatrix} \frac{1}{\lambda_1^{(1)}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\lambda_D^{(1)}} \end{pmatrix} x = \begin{pmatrix} \frac{x_1^2}{\lambda_1^{(1)}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{x_D^2}{\lambda_D^{(1)}} \end{pmatrix}$$

6.3 K-L变换方法Karhunen-Loeve transform

2. 运用DKLT消除两类问题的特征相关性方法

用U消除ω1分量的相关性



Step1： $U^T S_{W1} U = \Lambda_1$

Step 2: 做白化变换

$$B = U\Lambda_1^{-1/2}$$

$$\tilde{S}_{W_2} = B^{\mathrm{T}} S_{W_2} B$$

Step 3：求$\tilde{S}_{W_2}$特征值矢量矩阵V:

Step 4：得变换矩阵W

$$W = BV$$

$$z = W^T x = V^{\mathrm{T}} \Lambda_1^{-1/2} U^{\mathrm{T}} x \qquad x \in \omega_1 \text{ 或 } x \in \omega_2$$

6.3 K-L变换方法Karhunen-Loeve transform

3. 基于总的类内离差矩阵Sw的特征提取

先按S_W提供的信息产生相应的K-L坐标系，把原向量各分量**相关性消除**，得到在新坐标系中各分量离散程度。**再**对均值向量在新坐标系中的**分离程度**S_B做出判断

类间离散度尽量大

类内离散度尽量小

构造准则函数 $J(y_i) = \dfrac{u_i^{\mathrm{T}} S_B u_i}{u_i^{\mathrm{T}} S_W u_i} = \dfrac{u_i^{\mathrm{T}} S_B u_i}{\lambda_i}$

其中：$S_W \to \lambda_i \to u_i \quad (i = 1, 2, \cdots, D)$ $\qquad \lambda_i = u_i^{\mathrm{T}} S_W u_i$

该判据表征变换后的特征 $y_i = u_i^T X$ 的分类性能，y_i表示在新坐标轴u_i上的分量

S_b是类间离散度矩阵。可见$J(y_i)$是类间离散度与类内离散度在u_i坐标轴上的分量之比，$J(y_i)$越大，表明在新坐标系中该坐标轴包含越多的**可分性信息**。

排序：$J(y_1) \geq J(y_2) \geq \cdots \geq J(y_D)$ ，取前面 d 个较大的 $J(\cdot)$ 值对应的特征值 $u_i\ (i = 1,2,\cdots,d)$ 组成变换矩阵 W，有

$$Y = W^{\mathrm{T}} X$$

$$J(y_i)=\frac{u_i^{\mathrm{T}} S_B u_i}{\lambda_i}$$

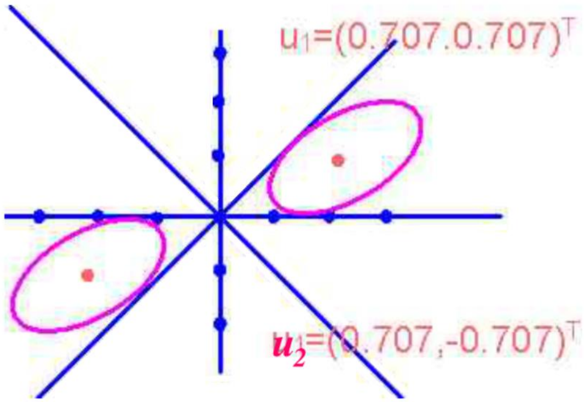
例题：设有两类问题，先验概率相等，样本均值向量分别为

$$\mu_1=[4,2]^T，\mu_2=[-4,-2]^T$$

协方差阵分别为

$$\Sigma_1=\begin{bmatrix}3 & 1\\ 1 & 3\end{bmatrix},\qquad \Sigma_2=\begin{bmatrix}4 & 2\\ 2 & 4\end{bmatrix}$$

要求把特征从2维压缩到1维。



解答：首先求S_W

$$S_W=\frac{1}{2}\Sigma_1+\frac{1}{2}\Sigma_2=\begin{bmatrix}3.5 & 1.5\\ 1.5 & 3.5\end{bmatrix}$$

求其特征值对角阵和特征向量矩阵为：

$$\Lambda=\begin{bmatrix}5 & 0\\ 0 & 2\end{bmatrix}\qquad \mathrm{U}=\begin{bmatrix}0.707 & 0.707\\ 0.707 & -0.707\end{bmatrix}$$

再求S_B

$$S_B=\sum_{i=1}^{c}P(\omega_i)(\mu_i-\mu)(\mu_i-\mu)^T=\begin{bmatrix}16 & 8\\ 8 & 4\end{bmatrix}$$

可计算得$J(y_1)$=3. 6，$J(y_2)$=1

因此 $u_1=[0.707,0.707]^T$ 作为一维特征空间的坐标轴，

6.3 K-L变换方法Karhunen-Loeve transform

4. 依据S_W，S_B作DKLT以降低特征维数的最优压缩方法

设Λ和U是对称正定阵S_W的特征值对角阵和特征矢量矩阵，做白化变换

$$V^T \Lambda^{-\frac{1}{2}} U^T S_B U \Lambda^{-\frac{1}{2}} V = \tilde{\Lambda}$$

正定矩阵和单位矩阵合同

存在正交阵V可使 $V^T \Lambda^{-\frac{1}{2}} U^T S_B U \Lambda^{-\frac{1}{2}} V = \tilde{\Lambda}$

其中$\tilde{\Lambda}$是白化变换后的总的类间离差阵$V^T \Lambda^{-\frac{1}{2}} U^T S_B U \Lambda^{-\frac{1}{2}} V = \tilde{\Lambda}$的特征值对角阵

由于S_B 的秩不大于c-1（对于C类问题，由于总的类间离差矩阵S_B的秩不大于c-1,故最多有c-1个非零特征值），所以$\tilde{S}_B$ 最多只有c-1个非0特征值，设非0特征值共有d个，用这d个非0特征值对应的特征矢量v_j(j=1,2,…,d)作变换矩阵，$V^T \Lambda^{-\frac{1}{2}} U^T S_B U \Lambda^{-\frac{1}{2}} V = \tilde{\Lambda}$。所得的d个分量含有原来D维模式的全部信息。

不损失信息而又达到最小维数的变换矩阵为：

$$W^T = V^T \Lambda^{-\frac{1}{2}} U^T \qquad V^T \Lambda^{-\frac{1}{2}} U^T S_B U \Lambda^{-\frac{1}{2}} V = \tilde{\Lambda}$$

基于总的类内离差矩阵SW的特征提取

$\tilde{S}_B$特征矢量阵

S_W特征值对角阵

S_W特征矢量阵

6.3 K-L变换方法Karhunen-Loeve transform

4. 依据S_W，S_B作DKLT以降低特征维数的最优压缩方法

步骤1：先用原坐标系中的S_W作为产生矩阵，进行K-L变换，消除原数据的相关性。所得到K-L坐标系中的新S_W'，是一个对角矩阵，其相应的K-L坐标系为U，由原S_W特征值对应的特征向量组成，Λ为对应特征向量矩阵。然后进一步变换，使S_W'变为单位阵。从原S_W到单位矩阵I有白化变换

$$\Lambda^{-\frac{1}{2}} U^T S_W U \Lambda^{-\frac{1}{2}} = I$$

经过白化变换B=$U\Lambda^{-\frac{1}{2}}$后得到类间离散矩阵 $\tilde{S}_B = \Lambda^{-\frac{1}{2}} U^T S_B U \Lambda^{-\frac{1}{2}}$=$B^T S_B B$

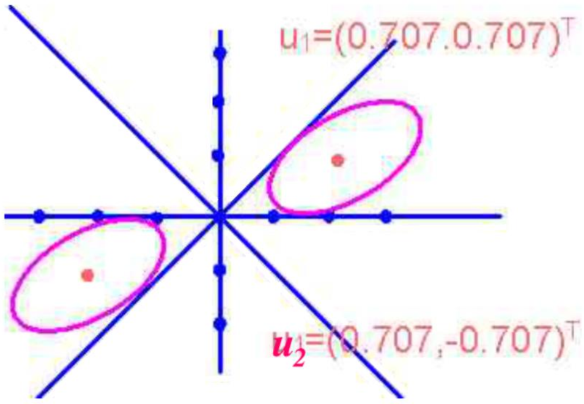
步骤2：以$\tilde{S}_B$作为产生矩阵，做第二次K-L变换，设$\tilde{S}_B$的秩为d（d≤c-1)，这d个特征值对应的特征向量矩阵为：

$$V = \begin{pmatrix} v_1 & v_2 & \cdots & v_d \end{pmatrix}$$

变换矩阵为： $W^T = V^T \Lambda^{-\frac{1}{2}} U^T$

例题：设有两类问题，先验概率相等，样本均值向量分别为

$\mu_1=[4,2]^T$，$\mu_2=[-4,-2]^T$



协方差阵分别为 $\Sigma_1=\begin{bmatrix}3 & 1\\ 1 & 3\end{bmatrix}$， $\Sigma_2=\begin{bmatrix}4 & 2\\ 2 & 4\end{bmatrix}$

要求把特征从2维压缩到1维，求依据S_W、S_B作DKL变换压缩方法的压缩结果。

解答：首先求S_W $S_W=\frac{1}{2}\Sigma_1+\frac{1}{2}\Sigma_2=\begin{bmatrix}3.5 & 1.5\\ 1.5 & 3.5\end{bmatrix}$

求其特征值对角阵和特征向量矩阵为： $\Lambda=\begin{bmatrix}5 & 0\\ 0 & 2\end{bmatrix}$ $U=\begin{bmatrix}0.707 & 0.707\\ 0.707 & -0.707\end{bmatrix}$

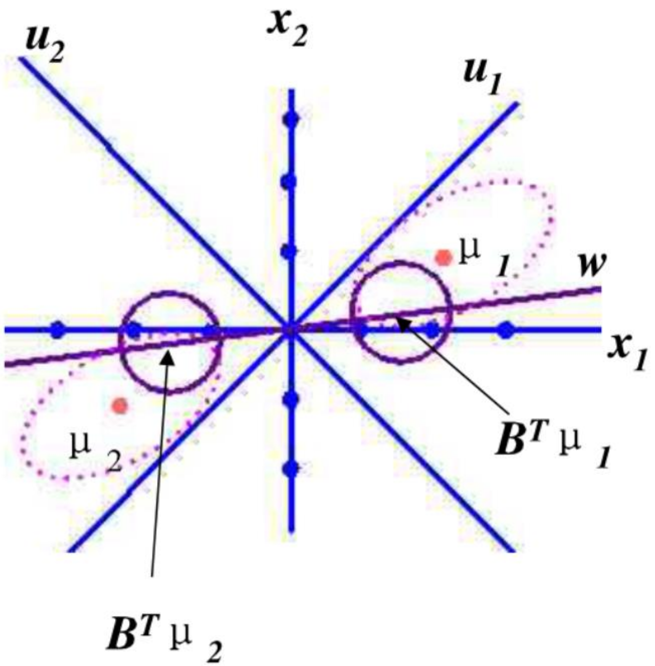
再求S_B

$$S_B=\sum_{i=1}^{c}P(\omega_i)(\mu_i-\mu)(\mu_i-\mu)^T=\begin{bmatrix}16 & 8\\ 8 & 4\end{bmatrix}$$

$$B = U\Lambda^{-1/2} = \begin{bmatrix} 0.707 & 0.707 \\ 0.707 & -0.707 \end{bmatrix}\begin{bmatrix} 0.447 & 0 \\ 0 & 0.707 \end{bmatrix} = \begin{bmatrix} 0.316 & 0.5 \\ 0.316 & -0.5 \end{bmatrix}$$

$$\tilde{S}_B = B^T S_B B = \begin{bmatrix} 3.6 & 1.897 \\ 1.897 & 1 \end{bmatrix}$$

$\tilde{S}_B$的特征值矩阵为： $\tilde{\Lambda} = \begin{bmatrix} 4.6 & 0 \\ 0 & 0 \end{bmatrix}$



非零特征值对应特征向量矩阵为：

$$v = \begin{bmatrix} 0.884 \\ 0.446 \end{bmatrix}$$

所以：

$$W = Bv = U\Lambda^{-\frac{1}{2}}v = \begin{bmatrix} 0.512 \\ 0.046 \end{bmatrix}$$

6.4 实例分析

PCA方法最著名的应用是在人脸识别中特征提取及数据维降低，输入200*200大小的人脸图像，提取灰度值作为原始特征，原始特征将达到40000维，这给后面分类器的处理将带来极大的难度。著名的人脸识别Eigenface算法就是采用PCA算法，用一个低维子空间描述人脸图像，同时又保存了识别所需要的信息。

- 人脸识别中，特征向量矩阵U称为特征脸（**Eigenface**）空间，其中的特征向量u_i进行量化后可以看出人脸轮廓。
- 设有N个人脸训练样本，每个样本由其像素灰度值组成一个向量x_i，则样本图像的像素点数即为x_i的维数**M=width*height**，由向量构成的训练样本集为 $\{X_1, X_2, \cdots, X_N\}$
- 该样本集的平均向量为 $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$，平均向量又叫平均脸。
- 样本集的协方差矩阵为 $C = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)\left(X_i - \overline{X}\right)^{\mathrm{T}}$

6.4 实例分析

- 求出协方差矩阵的特征向量$\boldsymbol{u_i}$和对应的特征值 λ_i，这些特征向量组成的矩阵$\boldsymbol{U}$就是人脸空间的正交基底，用它们的线性组合可以重构出样本中任意的人脸图像，并且图像信息集中在特征值大的特征向量中，即使丢弃特征值小的向量也不会影响图像质量。

- 将协方差矩阵的特征值按大到小排序：$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_d \geq \lambda_{d+1} \geq \cdots$

- 由大于λ_d的λ_i对应的特征向量构成主成分，主成分构成的变换矩阵为

$$U = \left(u_1, u_2, \cdots, u_d\right)$$

- 这样每一幅人脸图像都可以投影到 $U = \left(u_1, u_2, \cdots, u_d\right)$构成的特征脸子空间中，$\boldsymbol{U}$的维数为$\boldsymbol{M \times d}$。有了这样一个降维的子空间，任何一幅人脸图像都可以向其作投影 $y = U^T X$，即并获得一组坐标系数，即低维向量$\boldsymbol{y}$，维数$\boldsymbol{d \times 1}$，这组系数表明了图像在子空间的位置，从而可以作为人脸识别的依据。

6.4 实例分析

Note 1：

求的是相关矩阵 $R_X = E\left(XX^{'}\right)$ 的特征向量和特征值，这里怎么求的是协方差矩阵

$$C_X = \frac{1}{N}\sum_{i=1}^{N}\left(X-\overline{X}\right)\left(X-\overline{X}\right)^T$$

- 其实协方差矩阵也是：

$$C_X = E\left(\left(X-\overline{X}\right)\left(X-\overline{X}\right)^{'}\right)$$

- 可以看出其实用 $X-\overline{X}$，代替***X***就成了相关矩阵***R***，相当于原始样本向量都减去个平均向量，实质上还是一样的，协方差矩阵也是实对称矩阵。

Note 2：

在人脸识别过程中，对输入的一个测试样本*x*，求出它与平均脸的偏差 $X-\overline{X}$，则 $X-\overline{X}$ 在特征脸空间*U*的投影，可以表示为系数向量 *y*：

$$Y = U^{T}\left(X-\overline{X}\right)$$

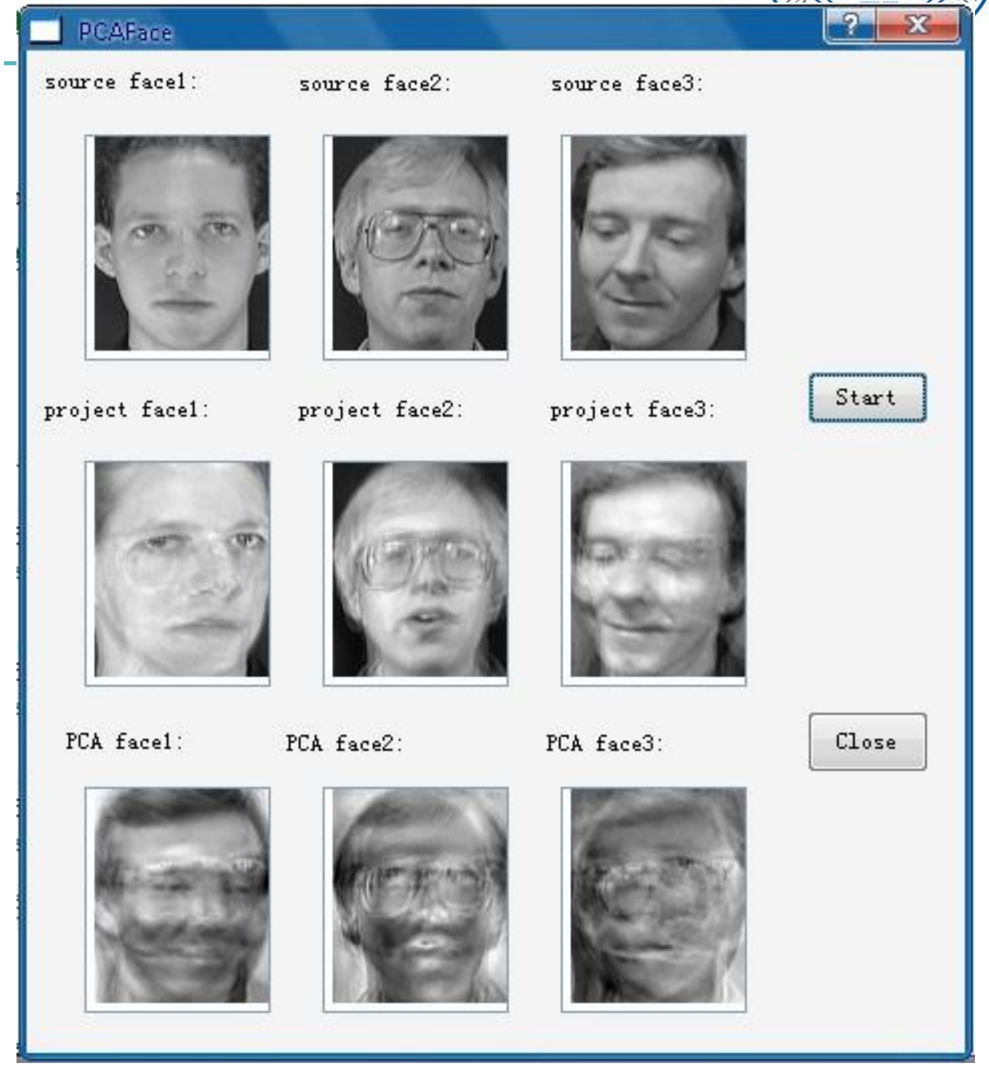
*U*的维数为$M\times d$，$X-\overline{X}$ 的维数为$M\times 1$，*y*的维数$d\times 1$。若*M*为200*200=40000维，取200个主成分，即200个特征向量，则最后投影的系数向量*y*维数降维到200维。

还有： $$X = UY + \overline{X}$$

这里的*x*就是根据投影系数向量*y*重构出的人脸图像，丢失了部分图像信息，但不会影响图像质量。

6.4 实例分析

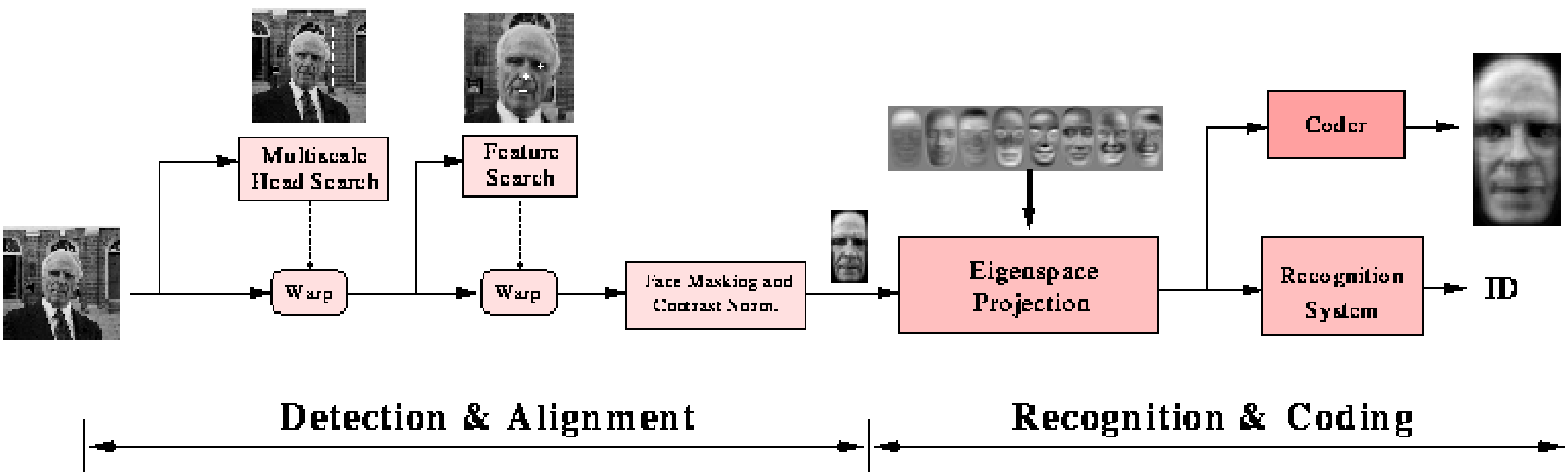
- 第一行的3张人脸分别为20张原图中的3张，这里取的是3个不同人的。
- 第二行中显示的3张人脸重构的人脸图像，可以看出由于只取了4个特征向量作为正交基底，因此重构后的人脸图像一些细节会丢失。如果增加保留的特征向量个数，则能较好的重构出人脸图像。
- 第三行的人脸图为取的原始数据协方差矩阵特征向量的最前面3个，因此这3个人脸为最具代表人脸特征的3个PCA人脸特征。



6.4 实例分析

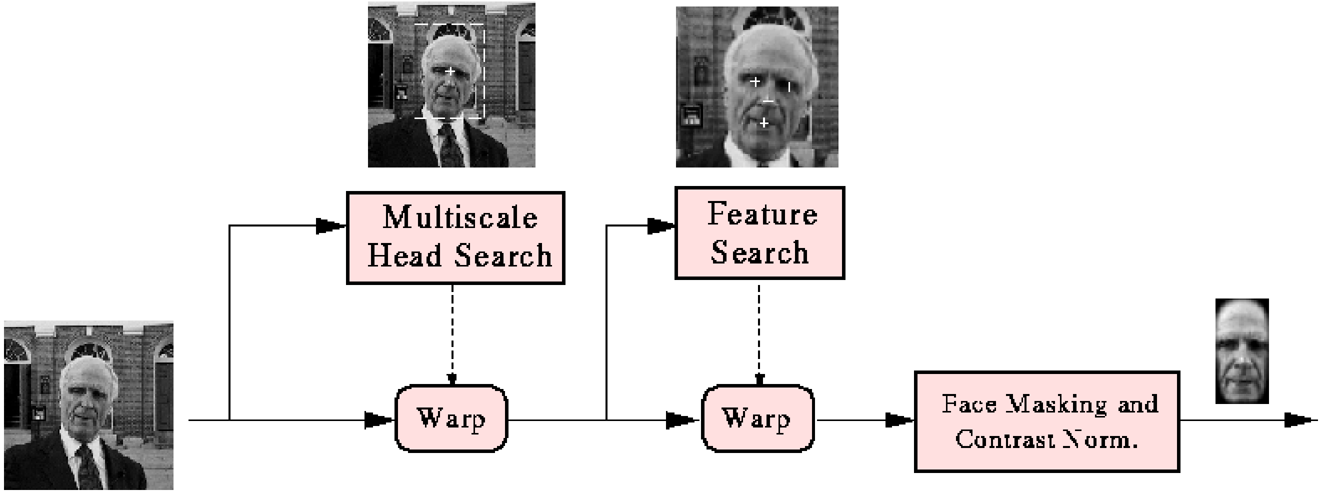
M. Turk & A. Pentland, Eigenfaces for recognitionI, Journal of Cognitive Neuroscience, vol.3, no.1, pp.71-86, 1991

The MIT system diagram

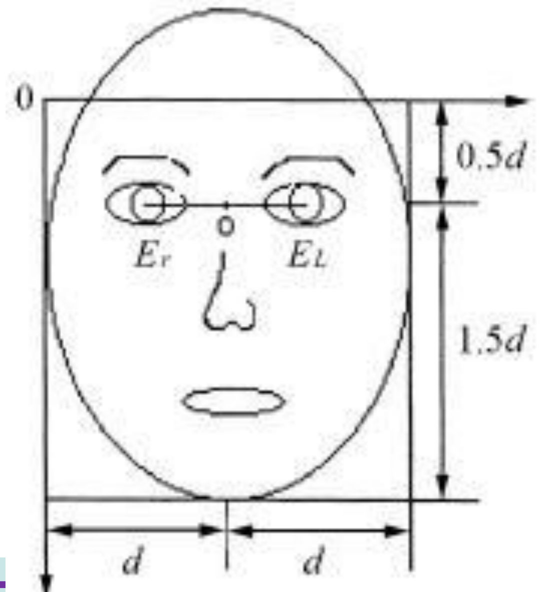


6.4 实例分析

预处理



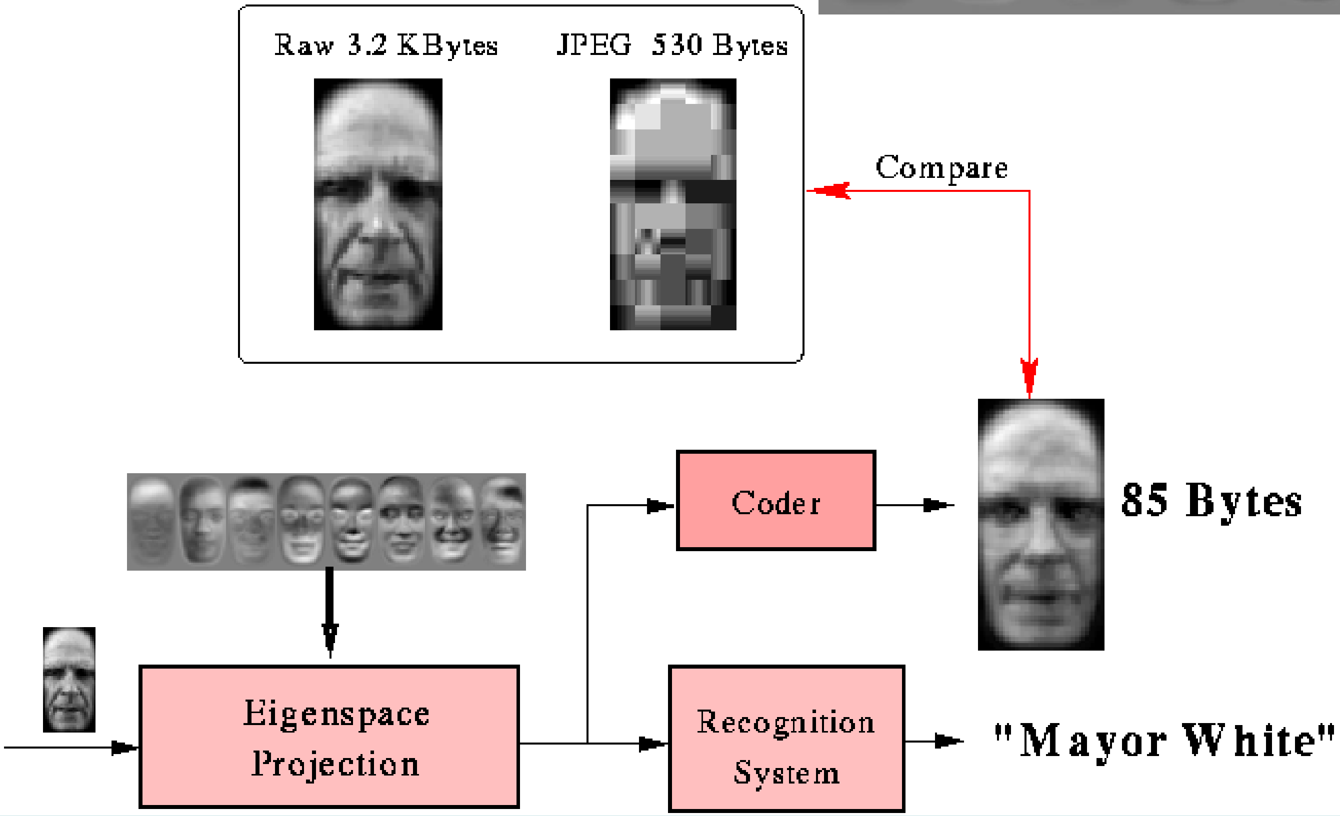
图像归一化和裁剪



6.4 实例分析

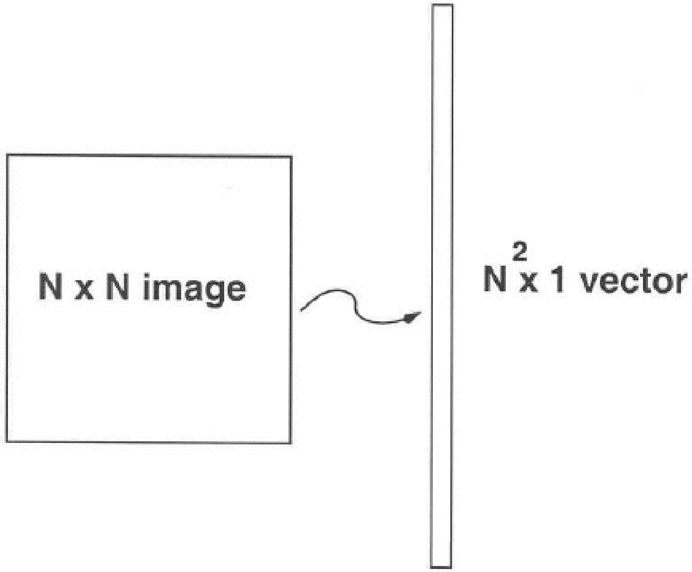
本征脸提取、表示和基于本征脸的分类

The first 8 normalized eigenfaces:



6.4 实例分析

方法



样本集 $\mathbf{x}_i \in R^{N^2}, i = 1, \cdots, M$

用**KL**变换（**PCA**）进行降维

总体散布矩阵 $\Sigma = \frac{1}{M}\sum_{i=0}^{M-1}(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T = \frac{1}{M}\mathbf{X}\mathbf{X}^T$

6.4 实例分析

$N^2 \times N^2$ 维矩阵，求其正交归一的本征向量，但计算困难。

解决办法：

考查M×M维矩阵 $\mathbf{R}=\mathbf{X}^T\mathbf{X}$，其特征方程是：

$$\mathbf{X}^T\mathbf{X}\mathbf{v}_i=\lambda_i\mathbf{v}_i$$

推导：

$$\mathbf{X}\mathbf{X}^T\mathbf{X}\mathbf{v}_i=\lambda_i\mathbf{X}\mathbf{v}_i$$

$$\mathbf{\Sigma}\mathbf{X}\mathbf{v}_i=\lambda_i\mathbf{X}\mathbf{v}_i$$

记 $\mathbf{u}_i=\mathbf{X}\mathbf{v}_i$，有 $\mathbf{\Sigma}\mathbf{u}_i=\lambda_i\mathbf{u}_i$

所以，矩阵 $\mathbf{X}^T\mathbf{X}$ 和 $\mathbf{X}\mathbf{X}^T$ 具有相同的本征值，而本征向量具有关系 $\mathbf{u}_i=\mathbf{X}\mathbf{v}_i$

6.4 实例分析

易求得， Σ 的归一化的本征向量是

$$\mathbf{u}_i = \frac{1}{\sqrt{\lambda_i}} \mathbf{X}\mathbf{v}_i, \quad i = 1,2,\cdots,M$$

注意，因为矩阵 Σ 的秩最多为**M**，所以最多只有**M**个本征值和本征向量。

每一个本征向量仍然是一个 N^2 维向量，即**N×N**维图像，仍然具有类似人脸的样子，因此被称作“本征脸”（**eigenfaces**）。

按照本征值从大到小排列， $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$

并从前向后取对应的本征脸，即构成对原图像的最佳的降维表示。

6.4 实例分析

原图像可以表示成本征脸的线性组合（在本征脸空间中的点）。

$$\mathbf{y}_{\mathbf{i}} = \mathbf{U}^T \mathbf{x}_i, i = 1, \cdots, M$$

$$y_{ij} = \mathbf{u}_j^T \mathbf{x}_i, i, j = 1, \cdots, M$$

6.4 实例分析

比如选取前k个本征向量，使

$$\sum_{i=0}^{k-1}\lambda_i \Big/ \sum_{i=0}^{M-1}\lambda_i \geq \alpha$$

比如 $\alpha = 99\%$，即可以保持原样本99%的信息。

对原图像的表示 $$\hat{\mathbf{x}}_i - \boldsymbol{\mu} = \sum_{j=1}^{k} y_{ij}\mathbf{u}_j$$

The original face and the recovered face

Ending