## 一、总体主成分

主成分，在代数学上是 $p$ 个随机变量 $X_1, X_2, \cdots, X_p$ 的一些特殊的线性组合。

设随机向量 $X = [X_1, X_2, \cdots, X_p]^T \in \mathbb{R}^p$，其协方差阵为 $\Sigma$，$\Sigma$ 的特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$

设系数向量 $a_i = [a_{i1}, a_{i2}, \cdots, a_{ip}]^T \in \mathbb{R}^p$，则有线性组合：

$$Y_i = a_i^T X = \sum_{j=1}^{p} a_{ij} X_j$$

易知，有

$$Var(Y_i) = a_i^T \Sigma a_i$$

$$Cov(Y_i, Y_k) = a_i^T \Sigma a_k$$

主成分是 $Y_i$ 中不相关的 $Y_1, Y_2, \cdots, Y_p$，同时要使得 $\forall\ Var(Y_i) = a_i^T \Sigma a_i$ 尽可能的大。

## 结论：总体主成分

设 $\Sigma$ 为随机向量 $X = [X_1, X_2, \cdots, X_p]^T \in \mathbb{R}^p$ 的协方差阵。$\Sigma$ 的特征值-特征向量为 $(\lambda_i, e_i)$ $i = 1, 2, \cdots, p$ 且 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ 则第 i 个主成分：

$$Y_i = e_i^T X \qquad i = 1, 2, \cdots, p$$

其中 $e_i = [e_{i1}, e_{i2}, \cdots, e_{ip}]^T \in \mathbb{R}^p$ $\quad Y_i \in \mathbb{R}$

此时有：

$$Var(Y_i) = e_i^T \Sigma e_i = \lambda_i$$

$$\boxed{Cov(Y_i, Y_k) = e_i^T \Sigma e_k = 0 \quad i \neq k}$$

证明：易知 $\max\limits_{a\neq 0} \dfrac{a^T\Sigma a}{a^T a} \overset{引理}{=} \lambda_1$（当 $a=e_1$ 取到）

加入 $e_1^T e_1 = 1$ 故有

$$Var(Y_1) = e_1^T \Sigma e_1 = \frac{e_1^T\Sigma e_1}{e_1^T e_1} = \lambda_1 = \max_{a\neq 0}\frac{a^T\Sigma a}{a^T a}$$

即 $Var(Y_1)$ 达到最大。

类似地，因为：$\max\limits_{a\perp e_1, e_2, \cdots, e_k} \dfrac{a^T\Sigma a}{a^T a} \overset{引理}{=} \lambda_{k+1}$

取 $a = e_{k+1}$ 使 $e_{k+1}^T e_i = 0 \quad (i=1,2,\cdots,k) \quad (k=1,2,\cdots,p-1)$

有 $e_{k+1}^T \Sigma e_{k+1} / e_{k+1}^T e_{k+1} = e_{k+1}^T \Sigma e_{k+1} = Var(Y_{k+1})$

而 $e_{k+1}^T (\Sigma e_{k+1}) = \lambda_{k+1} e_{k+1}^T e_{k+1} = \lambda_k$ 故

$$Var(Y_{k+1}) = \lambda_{k+1}$$

若 $\lambda_i$ 互不相同，有 $e_i^T e_k = 0 \quad i\neq k$ 则

$$Cov(Y_i, Y_k) = e_i^T \Sigma e_k = \lambda_k e_i^T e_k = 0$$

若 $\lambda_i$ 有相同，仍可以选出 $e_i^T e_k = 0$ 同理。

[引理] $B \in \mathbb{R}^{p\times p}$ 正定，特征值-特征向量 $(\lambda_i, e_i)$，不妨设 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ 则有：

(1) $\max\limits_{x\neq 0} \dfrac{x^T B x}{x^T x} = \lambda_1$ （当 $x = e_1$）

(2) $\min\limits_{x\neq 0} \dfrac{x^T B x}{x^T x} = \lambda_p$ （当 $x = e_p$）

(3) $\max\limits_{x\perp e_1,\cdots,e_k} \dfrac{x^T B x}{x^T x} = \lambda_{k+1}$ （当 $x = e_{k+1}$）

证明：

$B = P\Lambda P^T \quad \Lambda = diag(\lambda_i) \quad PP^T = I$ 记 $y = P^T x$

原式 $\frac{x^TBx}{x^Tx} = \frac{y^T\Lambda y}{y^Ty} = \sum \lambda_i y_i^2 / \sum y_i^2$

显然，上式为 $\lambda_i$ 的加权和 $\Rightarrow \in [\lambda_p, \lambda_1]$

故 (1) (2) 证毕，下面证明 (3)

已知 $x \perp e_1, e_2, \cdots, e_k$　$y = P^Tx \Rightarrow x = Py$，于是有 $Py$

与 $e_1, e_2, \cdots, e_k$ 正交，而 $x = Py = y_1e_1 + y_2e_2 + \cdots + y_pe_p$

有 $0 = e_i^Tx = e_i(\sum_{j=1}^{p} y_je_j)$ 对 $\forall\ i = 1,2,\cdots,k$ 由 $e_i$ 性质

有 $0 = 0 + y_i \Rightarrow y_i = 0$ $\forall\ i = 1,2,\cdots,k$ 则有

$$\frac{x^TBx}{x^Tx} = \sum \lambda_i y_i^2 / \sum y_i^2 = \sum_{i=k+1} \lambda_i y_i^2 / \sum_{i=k+1} y_i$$

同理可证：上式 $\leq \lambda_{k+1}$　$\square$

结论：

设随机向量 $X = [X_1, X_2, \cdots, X_p] \in \mathbb{R}^p$ 有协方差阵 $\Sigma$

其 eigenvalue - eigenvector pairs 为 $(\lambda_i, e_i)$ 且 $\lambda_1 \geq \cdots \geq \lambda_p$

令 $Y_i = e_i^TX$ 为主成分，则有：

$$\sum_{i=1}^{p} \sigma_{ii} = \sum_{i=1}^{p} Var(X_i) = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} Var(Y_i)$$

Proof：易知 $\sum \sigma_{ii} = tr(\Sigma)$ 对 $\Sigma$ 分解为 $P\Lambda P^T$

由迹的交换性：$\sum \sigma_{ii} = tr(\Sigma) = tr(P\Lambda P^T)$

$= tr(\Lambda P^TP) = tr(\Lambda) = \sum \lambda_i$

**定义**：总方差中属于第 k 个主成分的比例

$$\frac{\lambda_k}{\sum_{i=1}^{p}\lambda_i} \qquad k=1,2,\cdots,p$$

**结论**

若 $Y_i = e_i^T X$ 是从 $\Sigma$ 所得到的主成分，则

$$\rho_{Y_i, X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$$

其中 $\rho_{Y_i, X_k}$ $(i,k=1,2,\cdots,p)$ 为 $Y_i$ 与 $X_k$ 的相关系数，$e_{ik}$ 代表特征向量 $e_i$ 的第 k 个分量。

**Proof**：设 $a_k = [0,\cdots 0,1,0,\cdots,0]^T \in \mathbb{R}^p$ 第 k 个分量为 1，其他为 0。于是 $X_k = a_k^T X \in \mathbb{R}$

计算

$$\begin{aligned}
Cov(X_k, Y_i) &= Cov(a_k^T X, e_i^T X) \\
&= a_k^T \Sigma e_i \\
&= a_k^T \lambda_i e_i \\
&= \lambda_i e_{ik}
\end{aligned}$$

加之 $Var(Y_i) = \lambda_i$ $\quad Var(X_k) = \sigma_{kk}$ 代入有

$$\rho_{Y_i, X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$$

$\square$

## 椭球

若 $X \sim N_p(\mu, \Sigma)$ 则有椭球：

$$(X-\mu)^T \Sigma^{-1} (X-\mu) = c^2$$

记 $X' = X-\mu$ $\Sigma$ 分解为 $P\Lambda P^T$ 则

$$\Sigma^{-1} = (P\Lambda P^T)^{-1} = P\Lambda^{-1} P^T$$

于是 $X'^T \Sigma^{-1} X' = X'^T P \Lambda^{-1} P^T X'$

记 $P^T X' = y$ 有 $y^T \Lambda^{-1} y = \sum \frac{y_i^2}{\lambda_i}$

椭球形如

$$\sum_{i=1}^{p} \frac{y_i^2}{\lambda_i} = c^2$$

每个轴方向： $y_i = e_i^T X'$

## 从标准化变量得到主成分

$$Z_i = \frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}$$

利用矩阵记号：

$$Z = (V^{1/2})^{-1} (X-\mu)$$

其中 $V^{1/2} = \mathrm{diag}(\sqrt{\sigma_{ii}}) \in \mathbb{R}^{p\times p}$ $X, \mu \in \mathbb{R}^p$

易知

$$\mathrm{Cov}(Z) = (V^{1/2})^{-1} \Sigma (V^{1/2})^{-1} = \rho$$

其中 $\rho$ 为 $X$ 相关矩阵

## 结论：标准化主成分分析

有 $\mathrm{Cov}(Z)=\rho$ 的标准化变量 $Z^T=[Z_1,\cdots Z_p]$ 的第 i 个主成由

$$Y_i = e_i^T Z = e_i^T (V^{1/2})^{-1}(X-\mu)$$

给出，且有

$$\sum_{i=1}^{p} Var(Y_i) = \sum_{i=1}^{p} Var(Z_i) = p$$

$$\rho_{Y_i,Z_k} = e_{ik}\sqrt{\lambda_i} \qquad i,k=1,2,\cdots,p$$

此处 $(\lambda_i, e_i)$ 为 $\rho \in \mathbb{R}^{p\times p}$ 的特征对，且 $\lambda_i \ge \lambda_{i+1}$

此时总方差中属于第 k 个主成分的比例为

$$\lambda_k / p$$

Tip: 对任意随机向量 $X \in \mathbb{R}^p$ 有 2 种主成分分析方法：

① $\Sigma$ 分解 $(\tilde{\lambda}_i, \tilde{e}_i)$

$$Y_i = \tilde{e}_i^T X$$

② $\rho$ 分解 $(\hat{\lambda}_i, \hat{e}_i)$

$$Y_i = \hat{e}_i^T (V^{1/2})^{-1}(X-\mu)$$

选取原则：

① 各 $X_i$ 方差差异来自本身，使用 $\Sigma$

② 当差异来自量纲时，需要标准化，使用 $\rho$

# 特殊情况

①

$$\Sigma = \begin{bmatrix} \sigma_{11} & & & \\ & \sigma_{22} & & \\ & & \ddots & \\ & & & \sigma_{pp} \end{bmatrix}$$

则主成分 $Y_i = X_i$

②

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix} \quad 即 \quad \rho = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

则 $\lambda_1 = 1 + (p-1)\rho^2 \qquad \lambda_2 = \cdots = \lambda_p = 1-\rho$

$$e_1^T = \left[\frac{1}{\sqrt{p}}, \frac{1}{\sqrt{p}}, \cdots, \frac{1}{\sqrt{p}}\right]$$

$$e_i^T = \left[\frac{1}{\sqrt{(i-1)i}}, \cdots, \frac{1}{\sqrt{(i-1)i}}, \frac{-(i-1)}{\sqrt{(i-1)i}}, 0, \cdots, 0\right]$$

主成分 $Y_1 = e_1^T Z = \frac{1}{\sqrt{p}} \sum_{i=1}^{p} Z_i$

## 二、综合主成分的样本分量

设数据 $x_1, x_2, \cdots, x_n$ 来自 $(\mu, \Sigma)$ 的总体，这些数据：样本均值 $\bar{x}$，样本协方差 $S$，相关阵 $R$

### 样本主成分

对 $\forall\, x_j \in \mathbb{R}^p \quad (j=1,2,\cdots,n)$ 的线性组合：

$$a_k^T x = \sum_{i=1}^{p} a_{ki} x_{ji} \qquad j=1,2,\cdots,n$$

其中 $a_k = [a_{ki}]^T \in \mathbb{R}^p$.

于是，有样本均值 $a_k^T \bar{x}$，样本协方差 $a_k^T S a_k$

而 $\mathrm{Cov}(a_k^T x_j,\ a_l^T x_j) = a_k^T S a_l$

若 $S=\{s_{ik}\}$是特征值-特征向量对为$(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), \cdots, (\hat{\lambda}_p, \hat{e}_p)$的 $p\times p$ 样本协方差阵，则第 $i$ 个样本主成分由

x的分量 $x_i \in \mathbb{R}$

$$\hat{y}_i = \hat{e}_i' x = \hat{e}_{i1} x_1 + \hat{e}_{i2} x_2 + \cdots + \hat{e}_{ip} x_p, \qquad i=1,2,\cdots,p$$

给出，其中 $\hat{\lambda}_1 \geqslant \hat{\lambda}_2 \geqslant \cdots \geqslant \hat{\lambda}_p \geqslant 0$，而 $x$ 是变量 $X_1, X_2, \cdots, X_p$ 的任一观察，同时还有

$x \in \{x_1, x_2, \cdots, x_n\}$

$x_i \in \mathbb{R}^p$

$$\text{样本方差}(\hat{y}_k) = \hat{\lambda}_k, \quad k=1,2,\cdots,p \tag{8-20}$$

$$\text{样本协方差}(\hat{y}_i, \hat{y}_k) = 0, \quad i \neq k$$

此外，

$$\text{样本总方差} = \sum_{i=1}^{p} s_{ii} = \hat{\lambda}_1 + \hat{\lambda}_2 + \cdots + \hat{\lambda}_p$$

且

$$r_{\hat{y}_i, x_k} = \frac{\hat{e}_{ik}\sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}, \qquad i,k=1,2,\cdots,p$$

$X_1, X_2, \cdots, X_p \in \mathbb{R} \qquad x \in \mathbb{R}^p \qquad \hat{e}_i \in \mathbb{R}^p$

$S = \frac{1}{n-1}\sum (x_i - \bar{x})(x_i - \bar{x})^T \qquad x_i, \bar{x} \in \mathbb{R}^p$

样本主成分的标准化

## 样本主成分的标准化

一般说来，样本主成分在尺度改变时不是不变的(练习 8.2). 正如我们在总体成分处理中提到的，在不同尺度或在同一尺度但变化范围极其不同时所测量的变量，常常要标准化. 例如，利用下述构造完成标准化：

$$\mathbf{z}_j = \mathbf{D}^{-1/2}(\mathbf{x}_j - \bar{\mathbf{x}}) = \begin{bmatrix} \dfrac{x_{j1} - \bar{x}_1}{\sqrt{s_{11}}} \\ \dfrac{x_{j2} - \bar{x}_2}{\sqrt{s_{22}}} \\ \vdots \\ \dfrac{x_{jp} - \bar{x}_p}{\sqrt{s_{pp}}} \end{bmatrix} \qquad j = 1, 2, \cdots, n \tag{8-25}$$

$D^{1/2} = \operatorname{diag}(\sqrt{s_{ii}})$

将观测值标准化后的 $n \times p$ 数据矩阵

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1' \\ \mathbf{z}_2' \\ \vdots \\ \mathbf{z}_n' \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{bmatrix} \in \mathbb{R}^{n \times p}$$

$$= \begin{bmatrix} \dfrac{x_{11} - \bar{x}_1}{\sqrt{s_{11}}} & \dfrac{x_{12} - \bar{x}_2}{\sqrt{s_{22}}} & \cdots & \dfrac{x_{1p} - \bar{x}_p}{\sqrt{s_{pp}}} \\ \dfrac{x_{21} - \bar{x}_1}{\sqrt{s_{11}}} & \dfrac{x_{22} - \bar{x}_2}{\sqrt{s_{22}}} & \cdots & \dfrac{x_{2p} - \bar{x}_p}{\sqrt{s_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{x_{n1} - \bar{x}_1}{\sqrt{s_{11}}} & \dfrac{x_{n2} - \bar{x}_2}{\sqrt{s_{22}}} & \cdots & \dfrac{x_{np} - \bar{x}_p}{\sqrt{s_{pp}}} \end{bmatrix} \tag{8-26}$$

产生样本均值向量[见式(3-27)]

$$\frac{1}{n}(z_1+z_2+\cdots z_n)=0$$

$$\bar{\mathbf{z}} = \frac{1}{n}(\mathbf{1}'\mathbf{Z})' = \frac{1}{n}\mathbf{Z}'\mathbf{1} = \frac{1}{n}\begin{bmatrix} \sum_{j=1}^{n} \frac{x_{j1}-\bar{x}_1}{\sqrt{s_{11}}} \\ \sum_{j=1}^{n} \frac{x_{j2}-\bar{x}_2}{\sqrt{s_{22}}} \\ \vdots \\ \sum_{j=1}^{n} \frac{x_{jp}-\bar{x}_p}{\sqrt{s_{pp}}} \end{bmatrix} = \mathbf{0} \tag{8-27}$$

和样本协方差矩阵[见式(3-27)]

$$\begin{aligned} \mathbf{S}_z &= \frac{1}{n-1}\left(\mathbf{Z}-\frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{Z}\right)'\left(\mathbf{Z}-\frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{Z}\right) \\ &= \frac{1}{n-1}(\mathbf{Z}-\mathbf{1}\,\bar{\mathbf{z}}')'(\mathbf{Z}-\mathbf{1}\,\bar{\mathbf{z}}') \\ &= \frac{1}{n-1}\mathbf{Z}'\mathbf{Z} \\ &= \frac{1}{n-1}\begin{bmatrix} \frac{(n-1)s_{11}}{s_{11}} & \frac{(n-1)s_{12}}{\sqrt{s_{11}}\sqrt{s_{22}}} & \cdots & \frac{(n-1)s_{1p}}{\sqrt{s_{11}}\sqrt{s_{pp}}} \\ \frac{(n-1)s_{12}}{\sqrt{s_{11}}\sqrt{s_{22}}} & \frac{(n-1)s_{22}}{s_{22}} & \cdots & \frac{(n-1)s_{2p}}{\sqrt{s_{22}}\sqrt{s_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{(n-1)s_{1p}}{\sqrt{s_{11}}\sqrt{s_{pp}}} & \frac{(n-1)s_{2p}}{\sqrt{s_{22}}\sqrt{s_{pp}}} & \cdots & \frac{(n-1)s_{pp}}{s_{pp}} \end{bmatrix} = \mathbf{R} \end{aligned} \tag{8-28}$$

对角为1

结论

若 $z_1, z_2, \cdots, z_n$ 是协方差矩阵为 $\boldsymbol{R}$ 的标准化观测值，则第 $i$ 个样本主成分是

$$\hat{y}_i = \hat{\mathbf{e}}_i'\mathbf{z} = \hat{e}_{i1}z_1 + \hat{e}_{i2}z_2 + \cdots + \hat{e}_{ip}z_p, \qquad i = 1, 2, \cdots, p$$

其中 $(\hat{\lambda}_i, \hat{\mathbf{e}}_i)$ 是 $\boldsymbol{R}$ 的第 $i$ 个特征值-特征向量对，且 $\hat{\lambda}_1 \geqslant \hat{\lambda}_2 \geqslant \cdots \geqslant \hat{\lambda}_p \geqslant 0$. 另有

$$\text{样本方差}(\hat{y}_i) = \hat{\lambda}_i \qquad i = 1, 2, \cdots, p \tag{8-29}$$

$$\text{样本协方差}(\hat{y}_i, \hat{y}_k) = 0 \qquad i \neq k$$

此外，

$$(\text{标准化})\text{样本总方差} = \mathrm{tr}(\boldsymbol{R}) = p = \hat{\lambda}_1 + \hat{\lambda}_2 + \cdots + \hat{\lambda}_p$$

且

$$r_{\hat{y}_i, z_k} = \hat{e}_{ik}\sqrt{\hat{\lambda}_i}, \qquad i, k = 1, 2, \cdots, p$$

利用式(8-29)，我们看到，由第 $i$ 个样本主成分解释的样本总方差的比例是

$$\begin{pmatrix} \text{第 } i \text{ 个样本主成分} \\ \text{解释的(标准化)} \\ \text{样本方差所占的比例} \end{pmatrix} = \frac{\hat{\lambda}_i}{p} \qquad i = 1, 2, \cdots, p \tag{8-30}$$

## 三、 $p >> n$ 采用 SVD 分解

## 奇异值分解 SVD

对矩阵 $X \in \mathbb{R}^{n\times p}$ 分解为

$$X_{n\times p} = U \cdot \Lambda \cdot V^T$$

其中 $V \in \mathbb{R}^{p\times p}$ $\Lambda \in \mathbb{R}^{n\times p}$ $U, V$ 为正交阵 $UU^T = I$ $VV^T = I$

$U \in \mathbb{R}^{n\times n}$

于是

$$X^TX = (U\Lambda V^T)^T(U\Lambda V^T)$$
$$= V\Lambda^T U^T U \Lambda V^T$$
$$= V\Lambda^T\Lambda V^T \quad \in \mathbb{R}^{p\times p}$$
$$XX^T = U\Lambda\Lambda^T U^T \quad \in \mathbb{R}^{n\times n}$$

若 $u$ 为 $XX^T$ 的特征向量，特征值为 $w$ $(u \in \mathbb{R}^n)$

$$(XX^T)u = wu$$

左乘 $X^T$

$$X^T(XX^T)u = (X^TX)X^Tu = w\cdot(X^Tu)$$

即 $X^Tu$ 为 $X^TX$ 的关于 $w$ 的特征向量 $(X^Tu \in \mathbb{R}^p)$

## 过程 $X \in \mathbb{R}^{n\times p}$ $p >> n$

令 $\bar{\mu} = \frac{1}{n}1^TX$ $1 \in \mathbb{R}^n$ $\bar{\mu} \in \mathbb{R}^p$

令 $Y = \frac{1}{\sqrt{n-1}}\cdot(X - 1\cdot\bar{\mu}^T)$ $1 \in \mathbb{R}^{n\times 1}$ $\bar{\mu}^T \in \mathbb{R}^{1\times p}$

记 $YY^T$ 的特征向量为 $u \in \mathbb{R}^n$ $Y \in \mathbb{R}^{n\times p}$

有 $Y^T u = \Phi \in \mathbb{R}^p$ 为 $X^TX \in \mathbb{R}^{p\times p}$ 的特征向量

这样的 $u$ 共 $n$ 个 $(u_1, u_2, \cdots, u_n \in \mathbb{R}^n)$ 故

$\Phi$ 有 $n$ 个，$\Phi_1, \Phi_2, \cdots, \Phi_n \in \mathbb{R}^p$ 选前 $m$ 个构成

$$\hat{\Phi}_m = (\Phi_1, \Phi_2, \cdots, \Phi_m) \in \mathbb{R}^{p\times m}$$

对样本 $x \in \mathbb{R}^p$ 估计

$$\hat{y} = \hat{\Phi}_m^T (x - \bar{\mu}) \in \mathbb{R}^m$$

根据 $\hat{y}$ 分类，对于某个 $\hat{y}$，计算原 $x$：

$$x = \hat{\Phi}_m \cdot \hat{y} + \bar{\mu} \in \mathbb{R}^p$$