# 统计方法与机器学习

## 第四章：多重共线性 - 3

倪 葎

DaSE@ECNU
(lni@dase.ecnu.edu.cn)

# 目录

# 主成分回归

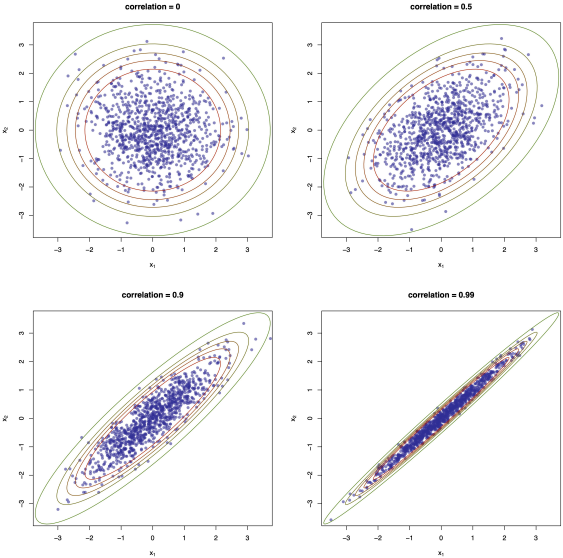
## 基本思想

$$\text{主成分回归} = \textbf{主成分}\text{分析} + \textbf{回归}\text{分析}$$

- 回归分析：研究因变量 $y$ 与自变量 $\boldsymbol{x} = (x_1, x_2, \cdots, x_p)'$ 之间的关系，即
$$y = \boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$
- 主成分分析：用 $k$ 个自变量的线性变换（主成分）代替原本的 $p$ 个自变量（$k < p$）；
- 关键问题：
  - 如何求主成分?
  - 如何利用主成分来估计回归参数 $\boldsymbol{\beta}$?
  - 主成分回归估计 $\hat{\boldsymbol{\beta}}_{\mathrm{PC}}$ 与最小二乘估计 $\hat{\boldsymbol{\beta}}$ 有什么关系?

# 主成分分析

## 动机


correlation = 0
correlation = 0.5
correlation = 0.9
correlation = 0.99
$x_1$
$x_2$
−3
−2
−1
0
1
2
3

# 主成分分析

## 动机

- 由于自变量个数太多，往往自变量之间存在着一定的相关性，因而使得所观测到的数据在一定程度上反映的信息有所重叠。

- 当自变量较多时，我们自然想到能用**较少**的综合变量代替原本多个自变量，而这几个综合变量有能够**尽可能多**地反映原本变量的信息，并且彼此之间互**不相关**。

# 主成分分析

## 基本想法

- 设 $\boldsymbol{x} = (x_1, x_2, \cdots, x_p)'$ 是 $p$ 维随机向量
  - 均值 $E(\boldsymbol{x}) = \boldsymbol{\mu}$;
  - 方差-协方差矩阵 $\mathrm{Var}(\boldsymbol{x}) = \boldsymbol{\Sigma}$.
- 考虑 $\boldsymbol{x}$ 的线性变换，即

$$\begin{cases} z_1 = \boldsymbol{a}_1'\boldsymbol{x} = a_{11}x_1 + a_{21}x_2 + \cdots + a_{p1}x_p \\ z_2 = \boldsymbol{a}_2'\boldsymbol{x} = a_{12}x_1 + a_{22}x_2 + \cdots + a_{p2}x_p \\ \vdots \\ z_p = \boldsymbol{a}_p'\boldsymbol{x} = a_{1p}x_1 + a_{2p}x_2 + \cdots + a_{pp}x_p \end{cases}$$

- 易知

$$\mathrm{Var}(z_i) = \boldsymbol{a}_i\boldsymbol{\Sigma}\boldsymbol{a}_i, \quad \mathrm{Cov}(z_i, z_j) = \boldsymbol{a}_i'\boldsymbol{\Sigma}\boldsymbol{a}_j$$

# 主成分分析

## 基本想法

- 假如，我们希望用一个变量 $z_1$ 来代替原来的 $p$ 个变量 $x_1, x_2, \cdots, x_p$，那么这个“新”变量 $z_1$ 需要满足
  - $z_1$ 能够尽可能多地反映原来 $p$ 个变量的信息。
- 问题：如何度量 $p$ 维变量的“信息量”?
- 通常，信息越不确定，信息量越大；
- 在统计学中，常常采用“方差”度量数据的不确定性；
- 于是，选取一个新变量 $z_1$。如果 $\mathrm{Var}(z_1)$ 越大，那么 $z_1$ 包含的信息量越多。

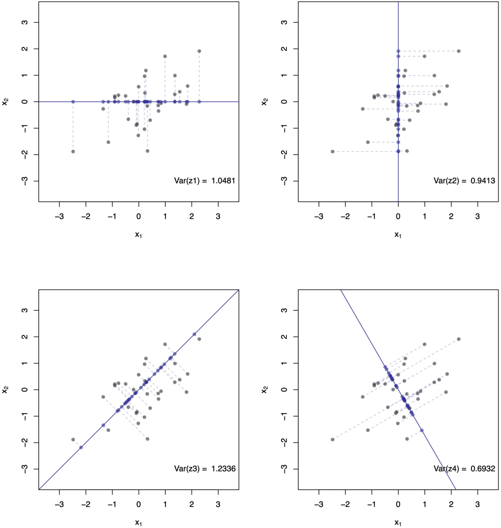
# 主成分分析

## 基本想法

- 假如，我们希望用一个变量 $z_1$ 来代替原来的 $p$ 个变量 $x_1, x_2, \cdots, x_p$，那么这个“新”变量 $z_1$ 需要满足
  - $z_1$ 能够尽可能多地反映原来 $p$ 个变量的信息。
- 问题：如何度量 $p$ 维变量的“信息量”?
- 通常，信息越不确定，信息量越大；
- 在统计学中，常常采用“方差”度量数据的不确定性；
- 于是，选取一个新变量 $z_1$。如果 $\mathrm{Var}(z_1)$ 越大，那么 $z_1$ 包含的信息量越多。

# 主成分分析

## 动机



# 主成分分析

## 基本想法

- 令 $z = \boldsymbol{a}'\boldsymbol{x} = a_1x_1 + a_2x_2 + \cdots + a_px_p$。
- 我们想要选取的变量 $z_1 = \boldsymbol{a}_1'\boldsymbol{x}$，使得

$$z_1 = \arg\max \mathrm{Var}(z) \quad \Leftrightarrow \quad \boldsymbol{a}_1 = \arg\max_{\boldsymbol{a}} \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}.$$

- 如果 $\|\boldsymbol{a}\| > \|\tilde{\boldsymbol{a}}\|$，那么

$$\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a} > \tilde{\boldsymbol{a}}'\boldsymbol{\Sigma}\tilde{\boldsymbol{a}}.$$

- 为了避免这个问题，我们需要对 $\boldsymbol{a}$ 做出一些限制。
- 最常用的限制是：$\boldsymbol{a}'\boldsymbol{a} = 1$。
- 若存在满足以上约束 $\boldsymbol{a}_1$，使得 $\mathrm{Var}(z_1)$ 达到最大，称 $z_1$ 为**第一主成分**。

## 基本想法

- 如果第一主成分不足以代表原来的 $p$ 个变量的绝大部分信息，我们需要进一步考虑 $\boldsymbol{x}$ 的第二个主成分 $z_2$。
- 为了有效地代表原始变量的信息，$z_1$ 已包含的信息并不希望体现在 $z_2$ 中.
- 从统计学的角度来看，要求

$$\mathrm{Cov}(z_2, z_1) = \boldsymbol{a}_2' \boldsymbol{\Sigma} \boldsymbol{a}_1 = 0$$

- 同样，在两个约束 $\boldsymbol{a}_2'\boldsymbol{a}_2 = 1$ 和 $\boldsymbol{a}_2'\boldsymbol{\Sigma}\boldsymbol{a}_1 = 0$ 下，当 $\mathrm{Var}(z_2)$ 达到最大时，确定 $a_2$。
- 类似地，可以求第三主成分、第四主成分、$\cdots$

**定义（主成分）**

设

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_p)'$$

为 $p$ 维随机向量且

$$\boldsymbol{a}_i = (a_{1i}, a_{2i}, \cdots, a_{pi})'$$

是个 $p$ 维常数向量。
如果 $z_i = \boldsymbol{a}_i'\boldsymbol{x}$ 是 $\boldsymbol{x}$ 的线性组合，且满足

- $\boldsymbol{a}_i'\boldsymbol{a}_i = 1$；
- 当 $i > 1$ 时，$\boldsymbol{a}_i'\boldsymbol{\Sigma}\boldsymbol{a}_j = 0$；
- $\mathrm{Var}(z_i) = \max\limits_{\boldsymbol{a}'\boldsymbol{a}=1, \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}_j=0, j=1,\cdots,i-1} \mathrm{Var}(\boldsymbol{a}'\boldsymbol{x})$。

那么，称 $z_i$ 为 $\boldsymbol{x}$ 的第 $i$ 个主成分。

## 主成分的求法：以第一主成分为例

- 设 $p$ 维随机向量 $\boldsymbol{x}$ 的均值 $E(\boldsymbol{x}) = 0$；
- $\boldsymbol{x}$ 的方差-协方差矩阵 $\mathrm{Var}(\boldsymbol{x}) = \boldsymbol{\Sigma}$.
- 问题：如何求第一主成分 $z_1 = \boldsymbol{a}_1'\boldsymbol{x}$?
- 求 $\boldsymbol{a}_1 = (a_{11}, a_{21}, \cdots, a_{p1})'$ 满足

$$\boldsymbol{a}_1 = \arg\max_{\boldsymbol{a}} \mathrm{Var}(\boldsymbol{a}'\boldsymbol{x}) \quad \text{s.t.} \quad \boldsymbol{a}'\boldsymbol{a} = 1$$

# 主成分分析

## 主成分的求法：以第一主成分为例

- 采用拉格朗日乘子法，令

$$l(\boldsymbol{a}) = \mathrm{Var}(\boldsymbol{a}'\boldsymbol{x}) - \lambda(\boldsymbol{a}'\boldsymbol{a} - 1) = \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a} - \lambda(\boldsymbol{a}'\boldsymbol{a} - 1)$$

- 于是有

$$\begin{cases} \dfrac{\partial l}{\partial \boldsymbol{a}} = 2(\boldsymbol{\Sigma} - \lambda \boldsymbol{I})\boldsymbol{a} = 0 \\ \dfrac{\partial l}{\partial \lambda} = \boldsymbol{a}'\boldsymbol{a} - 1 = 0. \end{cases}$$

- 因为 $\boldsymbol{a} \neq \mathbf{0}$，所以，$|\boldsymbol{\Sigma} - \lambda \boldsymbol{I}| = 0$，而且 $\boldsymbol{\Sigma}\boldsymbol{a} = \lambda\boldsymbol{a}$。
- 于是，求解第一主成分的问题等价于求 $\boldsymbol{\Sigma}$ 的特征值和特征向量问题。

# 主成分分析

**定理 3-4**

设

$$\boldsymbol{x} = (x_1, \cdots, x_p)'$$

是 $p$ 维随机向量，且 $\mathrm{Var}(\boldsymbol{x}) = \boldsymbol{\Sigma}$ 且满足

- $\boldsymbol{\Sigma}$ 的特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$;
- $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_p$ 为相应的单位正交特征向量.

则 $\boldsymbol{x}$ 的第 $i$ 主成分为

$$z_i = \boldsymbol{a}_i'\boldsymbol{x}, i = 1, 2, \cdots, p$$

- 可证明

$$\mathrm{Var}(\boldsymbol{z}) = \Lambda,$$

其中 $\boldsymbol{z} = (z_1, z_2, \cdots, z_p)'$ 且 $\Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_p\}$。

## 讨论

- 考虑以下情况:

$$\boldsymbol{\Sigma}_1=\begin{pmatrix}100 & 5\\ 5 & 1\end{pmatrix}\quad \text{且}\quad \boldsymbol{\Sigma}_2=\begin{pmatrix}1 & 0.5\\ 0.5 & 1\end{pmatrix}$$

- $\boldsymbol{\Sigma}_1$ 的特征值分别为 100.25，0.75；相应的特征向量为 $(0.9987, 0.0503)'$ 和 $(-0.0503, 0.9987)'$。
- $\boldsymbol{\Sigma}_1$ 的特征值分别为 1.5，0.5；相应的特征向量为 $(0.7071, 0.7071)'$ 和 $(-0.7071, 0.7071)'$。
- 根据方差-协方差矩阵 $\Sigma$ 来求主成分时，会是优先考虑方差大的变量，有时方差大是根据变量的量纲造成的。
- 对于方差差异大的变量，利用方差-协方差矩阵来求得主成分，可能得到不合理的结果。

# 主成分分析

## 讨论

- 考虑以下情况:

$$\mathbf{\Sigma}_1 = \begin{pmatrix} 100 & 5 \\ 5 & 1 \end{pmatrix} \quad 且 \quad \mathbf{\Sigma}_2 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

- $\mathbf{\Sigma}_1$ 的特征值分别为 100.25，0.75；相应的特征向量为 $(0.9987, 0.0503)'$ 和 $(-0.0503, 0.9987)'$。
- $\mathbf{\Sigma}_1$ 的特征值分别为 1.5，0.5；相应的特征向量为 $(0.7071, 0.7071)'$ 和 $(-0.7071, 0.7071)'$。
- 根据方差-协方差矩阵 $\mathbf{\Sigma}$ 来求主成分时，会是优先考虑方差大的变量，有时方差大是根据变量的量纲造成的。
- 对于方差差异大的变量，利用方差-协方差矩阵来求得主成分，可能得到不合理的结果。

# 主成分分析

## 讨论

- 为了消除由于量纲不同所带来的不合理结果，我们可以采用“标准化”后的方差-协方差矩阵来计算主成分。
- 令 $\mathrm{Corr}(\boldsymbol{x})$ 表示 $\boldsymbol{x}$ 的相关阵。
- 假定由 $\mathrm{Corr}(\boldsymbol{x})$ 所确定的主成分为 $\boldsymbol{z}^* = (z_1^*, \cdots, z_p^*)'$，则 $\boldsymbol{z}^*$ 的性质如下：
  - 主成分的协方差阵为

    $$\mathrm{Var}(\boldsymbol{z}^*) = \Lambda^* = \mathrm{diag}\{\lambda_1^*, \lambda_2^*, \cdots, \lambda_p^*\},$$

    其中，$\lambda_1^* \geq \cdots \geq \lambda_p^*$;
  - 特征值满足

    $$\sum_{i=1}^{p} \lambda_i^* = p$$

## 如何计算主成分?

- 样本数据（设计矩阵）为

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{pmatrix}.$$

## 如何计算主成分?

- 样本方差-协方差矩阵 $\boldsymbol{S}$，即

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})' \stackrel{\text{def}}{=} (s_{kl})_{p\times p}$$

其中,

$$\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i \stackrel{\text{def}}{=} (\bar{x}_1, \cdots, \bar{x}_p)'.$$

$$s_{kl} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ik} - \bar{x}_k)(x_{il} - \bar{x}_l).$$

# 主成分分析

## 如何计算主成分?

- 样本相关阵 $\boldsymbol{R}$ 为

$$\boldsymbol{R} = (r_{kl})_{p\times p}$$

其中

$$r_{kl} = \frac{s_{kl}}{\sqrt{s_{kk}s_{ll}}}$$

# 主成分分析

## 基本定义

- 假定标准化后的矩阵为

$$\boldsymbol{X}^* = \begin{pmatrix} x_{11}^* & x_{12}^* & \cdots & x_{1p}^* \\ x_{21}^* & x_{22}^* & \cdots & x_{2p}^* \\ \vdots & \vdots & & \vdots \\ x_{n1}^* & x_{n2}^* & \cdots & x_{np}^* \end{pmatrix}$$

- $x_{ij}$ 与 $x_{ij}^*$ 之间的关系

$$x_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{l_{jj}}}$$

其中，$l_{jj} = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2 = (n-1)s_{jj}$

## 基本定义

- 考虑 $(\boldsymbol{X}^*)'\boldsymbol{X}^*$ 中的每一个元素

$$\begin{aligned}
\sum_{i=1}^{n}(x_{ik}^*)(x_{il}^*) &= \sum_{i=1}^{n}\frac{x_{ik}-\bar{x}_k}{\sqrt{l_{kk}}}\times\frac{x_{il}-\bar{x}_l}{\sqrt{l_{ll}}} \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_{ik}-\bar{x}_k)(x_{il}-\bar{x}_l)}{\sqrt{s_{kk}s_{ll}}} \\
&= r_{kl}
\end{aligned}$$

- $(\boldsymbol{X}^*)'\boldsymbol{X}^*$ 等价于原始设计矩阵 $\boldsymbol{X}$ 的样本相关阵 $\boldsymbol{R}$。

## 动机

- 假定设计矩阵 $\boldsymbol{X}$ 已标准化。假设 $\boldsymbol{X}'\boldsymbol{X}$ 的
  - 特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$;
  - 其相应单位正交化后的特征向量为 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_p$.
- 我们令
  - $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_p\}$;
  - $\boldsymbol{V}' = (\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_p)$.
- 易知,
  - $\boldsymbol{V}'$ 是由相互正交的列向量组成，且 $\boldsymbol{V}\boldsymbol{V}' = \boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}$;
  - 特征值分解为 $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{V}'\boldsymbol{\Lambda}\boldsymbol{V}$ 即

$$\boldsymbol{V}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{V}' = (\boldsymbol{X}\boldsymbol{V}')'(\boldsymbol{X}\boldsymbol{V}') = \boldsymbol{\Lambda}.$$

# 主成分回归

## 动机

- 令 $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{V}'$，那么

$$\boldsymbol{Z}'\boldsymbol{Z} = \boldsymbol{\Lambda}.$$

- 在矩阵 $\boldsymbol{Z}_{n\times p} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_p)$ 中第 $j$ 个列向量 $\boldsymbol{z}_j$ 为 $n \times 1$ 向量，且满足

$$\boldsymbol{z}_j \quad = \quad \boldsymbol{X}\boldsymbol{v}_j = (z_{1j}, z_{2j}, \cdots, z_{nj})'$$

- 第 $j$ 个主成分为

$$\boldsymbol{z}_j = v_{1j}\boldsymbol{x}_1 + v_{2j}\boldsymbol{x}_2 + \cdots + v_{pj}\boldsymbol{x}_p$$

# 主成分回归

## 动机

- 由于 $\boldsymbol{X}$ 是已经标准化，即

$$\sum_{i=1}^{n} x_{ij} = 0, \quad \sum_{i=1}^{n} x_{ij}^2 = 1, \quad j = 1, 2, \cdots, p.$$

- 而且 $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{V}'$ 和 $\boldsymbol{Z}'\boldsymbol{Z} = \boldsymbol{\Lambda}$，我们有

$$\sum_{i=1}^{n} z_{ij} = 0, \quad \sum_{i=1}^{n} z_{ij}^2 = \lambda_j, \quad \sum_{i=1}^{n} z_{ij} z_{ik} = 0 \ (j \neq k).$$

- 这表明了
    - 矩阵 $\boldsymbol{Z}$ 的各列之间正交；
    - 当 $\lambda_j \approx 0$ 时，$z_{1j}, z_{2j}, \cdots, z_{nj}$ 均近似为 0；

# 主成分回归

## 主成分估计的定义

- 当 $|\boldsymbol{X}'\boldsymbol{X}| \approx 0$ 时，存在一个 $k$，使得 $\lambda_{k+1}, \cdots, \lambda_p$ 均近似为 0。因此，$\boldsymbol{z}_{k+1}, \cdots, \boldsymbol{z}_p$ 近似为 0。
- 线性回归模型简化为

$$\boldsymbol{y} \quad = \quad \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \boldsymbol{X}\boldsymbol{V}'\boldsymbol{V}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \boldsymbol{Z}_{n\times p}\boldsymbol{\alpha}_{p\times 1} + \boldsymbol{\varepsilon}$$

- 我们将矩阵 $\boldsymbol{Z}$ 和向量 $\boldsymbol{\alpha}$ 按以下方式拆分

$$\begin{aligned}
\boldsymbol{\alpha} &= (\alpha_1, \cdots, \alpha_k, \alpha_{k+1}, \cdots, \alpha_p)' = (\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2')', \\
\boldsymbol{Z} &= (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_k, \boldsymbol{z}_{k+1}, \cdots, \boldsymbol{z}_p) = (\boldsymbol{Z}_1, \boldsymbol{Z}_2).
\end{aligned}$$

- 回归模型也可以写为

$$\boldsymbol{y} = \boldsymbol{Z}_1\boldsymbol{\alpha}_1 + \boldsymbol{Z}_2\boldsymbol{\alpha}_2 + \boldsymbol{\varepsilon} = \boldsymbol{Z}_1\boldsymbol{\alpha}_1 + \boldsymbol{\varepsilon}$$

## 主成分估计的定义

- 线性回归模型

$$\boldsymbol{y} = \boldsymbol{Z}_1 \boldsymbol{\alpha}_1 + \boldsymbol{\varepsilon},$$

- $\boldsymbol{\alpha}_1$ 的最小二乘估计为

$$\hat{\boldsymbol{\alpha}}_1 = (\boldsymbol{Z}_1' \boldsymbol{Z}_1)^{-1} \boldsymbol{Z}_1' \boldsymbol{y} = \boldsymbol{\Lambda}_{\boldsymbol{1}}^{-1} \boldsymbol{Z}_1' \boldsymbol{y}$$

  其中, $\boldsymbol{\Lambda}_1 = \mathrm{diag}\{\lambda_1, \cdots, \lambda_k\}$, $\boldsymbol{\Lambda}_2 = \mathrm{diag}\{\lambda_{k+1}, \cdots, \lambda_p\}$.

- 那么，$\boldsymbol{\Lambda}$ 可以拆分为 $\boldsymbol{\Lambda}_1$ 和 $\boldsymbol{\Lambda}_2$，即

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_2 \end{pmatrix}$$

- 相应地，我们也可以将 $\boldsymbol{V}'$ 做拆分，即 $\boldsymbol{V}' = (\boldsymbol{V}_1', \boldsymbol{V}_2')$.

## 主成分估计的定义

- 根据

$$\boldsymbol{\beta} = \boldsymbol{V}'\boldsymbol{\alpha},$$

- 回归参数 $\boldsymbol{\beta}$ 的估计为

$$\hat{\boldsymbol{\beta}}_{\mathrm{PC}} = \boldsymbol{V}'\begin{pmatrix} \hat{\boldsymbol{\alpha}}_1 \\ \mathbf{0} \end{pmatrix} = \boldsymbol{V}_1'\hat{\boldsymbol{\alpha}}_1 = \boldsymbol{V}_1'\boldsymbol{\Lambda}_1^{-1}\boldsymbol{Z}_1'\boldsymbol{y}$$

- 称 $\hat{\boldsymbol{\beta}}_{\mathrm{PC}}$ 为 $\boldsymbol{\beta}$ 的**主成分估计**。

# 主成分回归

## 性质 1

- 主成分估计与最小二乘估计有什么关系吗?
- 主成分估计可写为最小二乘估计的线性变换，即

$$
\begin{aligned}
\hat{\beta}_{\mathrm{PC}} &= V_1'\Lambda_1^{-1}Z_1'y \\
&= V_1'\Lambda_1^{-1}V_1X'y \\
&= V_1'\Lambda_1^{-1}V_1X'X(X'X)^{-1}X'y \\
&= V_1'\Lambda_1^{-1}V_1X'X\hat{\beta} \\
&= V_1'\Lambda_1^{-1}V_1V'\Lambda V\hat{\beta} \\
&= V_1'\Lambda_1^{-1}V_1(V_1',V_2')\begin{pmatrix}\Lambda_1 & 0\\ 0 & \Lambda_2\end{pmatrix}\begin{pmatrix}V_1\\ V_2\end{pmatrix}\hat{\beta} \\
&= V_1'V_1\hat{\beta}.
\end{aligned}
$$

# 主成分回归

## 性质 1

- 主成分估计与最小二乘估计有什么关系吗?
- 主成分估计可写为最小二乘估计的线性变换，即

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{\mathrm{PC}} &= \boldsymbol{V}_1'\boldsymbol{\Lambda}_1^{-1}\boldsymbol{Z}_1'\boldsymbol{y} \\
&= \boldsymbol{V}_1'\boldsymbol{\Lambda}_1^{-1}\boldsymbol{V}_1\boldsymbol{X}'\boldsymbol{y} \\
&= \boldsymbol{V}_1'\boldsymbol{\Lambda}_1^{-1}\boldsymbol{V}_1\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\
&= \boldsymbol{V}_1'\boldsymbol{\Lambda}_1^{-1}\boldsymbol{V}_1\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} \\
&= \boldsymbol{V}_1'\boldsymbol{\Lambda}_1^{-1}\boldsymbol{V}_1\boldsymbol{V}'\boldsymbol{\Lambda}\boldsymbol{V}\hat{\boldsymbol{\beta}} \\
&= \boldsymbol{V}_1'\boldsymbol{\Lambda}_1^{-1}\boldsymbol{V}_1(\boldsymbol{V}_1', \boldsymbol{V}_2')\begin{pmatrix}\boldsymbol{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{V}_1 \\ \boldsymbol{V}_2\end{pmatrix}\hat{\boldsymbol{\beta}} \\
&= \boldsymbol{V}_1'\boldsymbol{V}_1\hat{\boldsymbol{\beta}}.
\end{aligned}
$$

# 主成分回归

## 性质 2

- 主成分估计是无偏估计吗?
- 主成分估计的期望为

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}_{\mathrm{PC}}) &= E(\boldsymbol{V}_1'\boldsymbol{V}_1\hat{\boldsymbol{\beta}}) \\ &= \boldsymbol{V}_1'\boldsymbol{V}_1 E(\hat{\boldsymbol{\beta}}) \\ &= \boldsymbol{V}_1'\boldsymbol{V}_1\boldsymbol{\beta} \end{aligned}$$

- 由于 $\boldsymbol{I}_p = \boldsymbol{V}'\boldsymbol{V} = \boldsymbol{V}_1'\boldsymbol{V}_1 + \boldsymbol{V}_2'\boldsymbol{V}_2$,那么 $\boldsymbol{V}_1'\boldsymbol{V}_1 = \boldsymbol{I}_p - \boldsymbol{V}_2'\boldsymbol{V}_2$.
- 当 $k < p$ 时，$\boldsymbol{V}_1'\boldsymbol{V}_1\boldsymbol{\beta} = (\boldsymbol{I} - \boldsymbol{V}_2'\boldsymbol{V}_2)\boldsymbol{\beta} \neq \boldsymbol{\beta}$。
- 此时，主成分估计是有偏估计。

## 性质 3

- 虽然主成分估计是有偏估计，但是在均方误差的意义下，主成分估计是否优于最小二乘估计?

**定理 3-5**

当设计矩阵 $\boldsymbol{X}$ 存在多重共线性，选择合适的 $k$，可使得

$$\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{\mathrm{PC}}) < \mathrm{MSE}(\hat{\boldsymbol{\beta}})$$

# 主成分回归

## 性质 3

**证明：** 由于

$$\hat{\boldsymbol{\beta}}_{\mathrm{PC}} = \boldsymbol{V}' \begin{pmatrix} \hat{\boldsymbol{\alpha}}_1 \\ \boldsymbol{0} \end{pmatrix}$$

我们有

$$\begin{aligned}
\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{\mathrm{PC}}) &= E(\hat{\boldsymbol{\beta}}_{\mathrm{PC}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}_{\mathrm{PC}} - \boldsymbol{\beta}) \\
&= E\left(\boldsymbol{V}' \begin{pmatrix} \hat{\boldsymbol{\alpha}}_1 \\ \boldsymbol{0} \end{pmatrix} - \boldsymbol{V}'\boldsymbol{\alpha}\right)' \left(\boldsymbol{V}' \begin{pmatrix} \hat{\boldsymbol{\alpha}}_1 \\ \boldsymbol{0} \end{pmatrix} - \boldsymbol{V}'\boldsymbol{\alpha}\right) \\
&= E\left(\begin{pmatrix} \hat{\boldsymbol{\alpha}}_1 \\ \boldsymbol{0} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{pmatrix}\right)' \boldsymbol{V}\boldsymbol{V}' \left(\begin{pmatrix} \hat{\boldsymbol{\alpha}}_1 \\ \boldsymbol{0} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{pmatrix}\right) \\
&= E\left(\begin{pmatrix} \hat{\boldsymbol{\alpha}}_1 \\ \boldsymbol{0} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{pmatrix}\right)' \left(\begin{pmatrix} \hat{\boldsymbol{\alpha}}_1 \\ \boldsymbol{0} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{pmatrix}\right)
\end{aligned}$$

## 性质 3

**证明:**

$$
\begin{aligned}
\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{\mathrm{PC}}) &= E(\hat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)'(\hat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1) + \|\boldsymbol{\alpha}_2\|^2 \\
&= E(\boldsymbol{\varepsilon}' \boldsymbol{Z}_1 \boldsymbol{\Lambda}_1^{-2} \boldsymbol{Z}_1' \boldsymbol{\epsilon}) + \|\boldsymbol{\alpha}_2\|^2 \\
&= \sigma^2 \mathrm{tr}\left(\boldsymbol{\Lambda}_1^{-1}\right) + \|\boldsymbol{\alpha}_2\|^2 \\
&= \sigma^2 \sum_{j=1}^{k} \lambda_j^{-1} + \sum_{j=k+1}^{p} \alpha_j^2 \\
&= \mathrm{MSE}(\hat{\boldsymbol{\beta}}) + \left( \sum_{j=k+1}^{p} \alpha_j^2 - \sigma^2 \sum_{j=k+1}^{p} \lambda_j^{-1} \right)
\end{aligned}
$$

# 主成分回归

## 性质 3

**证明：** 我们已证明了

$$\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{\mathrm{PC}}) = \mathrm{MSE}(\hat{\boldsymbol{\beta}}) + \left(\sum_{j=k+1}^{p} \alpha_j^2 - \sigma^2 \sum_{j=k+1}^{p} \lambda_j^{-1}\right)$$

由于设计矩阵存在多重共线性，因此有一部分特征值 $\lambda_j$ 非常接近于零。
不妨设后 $p-k$ 个特征值接近于零，则 $\sum_{j=k+1}^{p} \lambda_j^{-1}$ 将会很大，这导致了第二项为负。所以，

$$\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{\mathrm{PC}}) < \mathrm{MSE}(\hat{\boldsymbol{\beta}})$$

# 主成分回归

## 性质 4

- 主成分估计的长度 vs 最小二乘估计的长度
- 主成分估计的模的平方为

$$
\begin{aligned}
\|\hat{\boldsymbol{\beta}}_{\mathrm{PC}}\|^2 &= (\hat{\boldsymbol{\beta}}_{\mathrm{PC}})'\hat{\boldsymbol{\beta}}_{\mathrm{PC}} \\
&= (\boldsymbol{V}_1'\boldsymbol{V}_1\hat{\boldsymbol{\beta}})'(\boldsymbol{V}_1'\boldsymbol{V}_1\hat{\boldsymbol{\beta}}) \\
&= \hat{\boldsymbol{\beta}}'\boldsymbol{V}_1'\boldsymbol{V}_1\boldsymbol{V}_1'\boldsymbol{V}_1\hat{\boldsymbol{\beta}} \\
&= \hat{\boldsymbol{\beta}}'\boldsymbol{V}_1'\boldsymbol{V}_1\hat{\boldsymbol{\beta}} \\
&\leq \hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}} \\
&= \|\hat{\boldsymbol{\beta}}\|^2
\end{aligned}
$$

- 所以，$\|\hat{\boldsymbol{\beta}}_{\mathrm{PC}}\| \leq \|\hat{\boldsymbol{\beta}}\|$ 主成分估计是压缩估计。

## 如何选择主成分的个数?

- **保留特征值比重大的主成分**
  - 由于 $\sum_{j=1}^{p} \lambda_j = p$, 通常称 $\frac{\lambda_j}{p}$ 为第 $j$ 个主成分 $z_j$ 的贡献率。而 $\sum_{j=1}^{k} \frac{\lambda_j}{p}$ 为前 $k$ 个主成分的累积贡献率。
  - 具体方案：给定一个定值 $c_{pc}(0 < c_{pc} < 1)$，如果存在 $k$，使得
  $$\sum_{j=1}^{k-1} \frac{\lambda_j}{p} < c_{\rm pc}, \quad \sum_{j=1}^{k} \frac{\lambda_j}{p} \geq c_{\rm pc}.$$
  由此选取 $k$。
  - 通常，$c_{\rm pc} = 70\%, 75\%$ 或 $80\%$。

## 主成分回归

如何选择主成分的个数?

- **删除特征值接近于令的主成分**
  - 具体方案：给定一个定值 $c_0$，如果

$$\lambda_k \geq c_0, \lambda_{k+1} < c_0$$

由此选取 $k$。

- **均方误差确定 $k$**
  - 由于 $\sum_{j=1}^{p} \lambda_j^{-1}$ 与估计的均方误差有关，我们并不希望这个值太大，我们可以选取 $k$ 满足

$$\sum_{j=1}^{k} \lambda_j^{-1} \leq 5k.$$