回归分析

特征选择

线性模型特征选择（Linear model variable selection）

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon.$$

子集**选择**（subset selection）：
- Adjusted R^2
- Mellow Cp
- AIC、BIC
- Validation selection
- N-fold validation selection（k折交叉验证）

特征**缩**减（Shrinkage method）：
- Lasso regression
- Ridge regression

降**维**（dimension deduction）：
- PCR
- PLS 偏最小二乘法

逐步回归（Stepwise Regression ）

- $\{x_1, x_2, x_3\}$

- 全选择
    1. $\{x_1\}$, $\{x_2\}$, $\{x_3\}$ 建立回归模型 找出RSS最小。
    2. $\{x_1, x_2\}$, $\{x_1, x_3\}$, $\{x_2, x_3\}$建立回归模型找出最小RSS的。
    3. $\{x_1, x_2, x_3\}$建立回归模型。

向前**选择**（Forward selection **）**

- $\{x_1, x_2, x_3\}$

- 向前选择

  1. $\{x_1\}$, $\{x_2\}$, $\{x_3\}$ 建立回归模型 找出RSS最小。

     比如选中的是$\{x_1\}$
  2. $\{x_1, x_2\}$, $\{x_1, x_3\}$ 建立回归模型找出最小RSS的。
  3. $\{x_1, x_2, x_3\}$建立回归模型。

向后**选择**（Backward selection ）

- $\{x_1, x_2, x_3\}$

- 向前选择
  1. $\{x_1, x_2, x_3\}$建立回归模型。
  2. $\{x_1, x_2\}, \{x_1, x_3\}, \{x_2, x_3\}$建立回归模型找出最小RSS的。
     假如是$\{x_1, x_3\}$最小
  3. $\{x_1\}, \{x_3\}$建立回归模型找出最小RSS的。

逐步回归（Stepwise Regression ）

- 向前向后选则进行了$1+\frac{p(p+1)}{2}$次回归
- 全选择进行了2^{p}次回归

子集选择

1. 逐步回归 $\{x_1, x_2, x_3\}, \{x_2, x_3\}, \{x_3\}$

2. 从逐步回归中选出最好的模型
   - Adjusted R^2
   - Mellow Cp
   - AIC、BIC
   - Validation selection
   - N-fold validation selection 交叉验证

特征选择

从逐步回归中选出最好的模型

- Adjusted R^2

$$R^2_{adjusted} = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

where

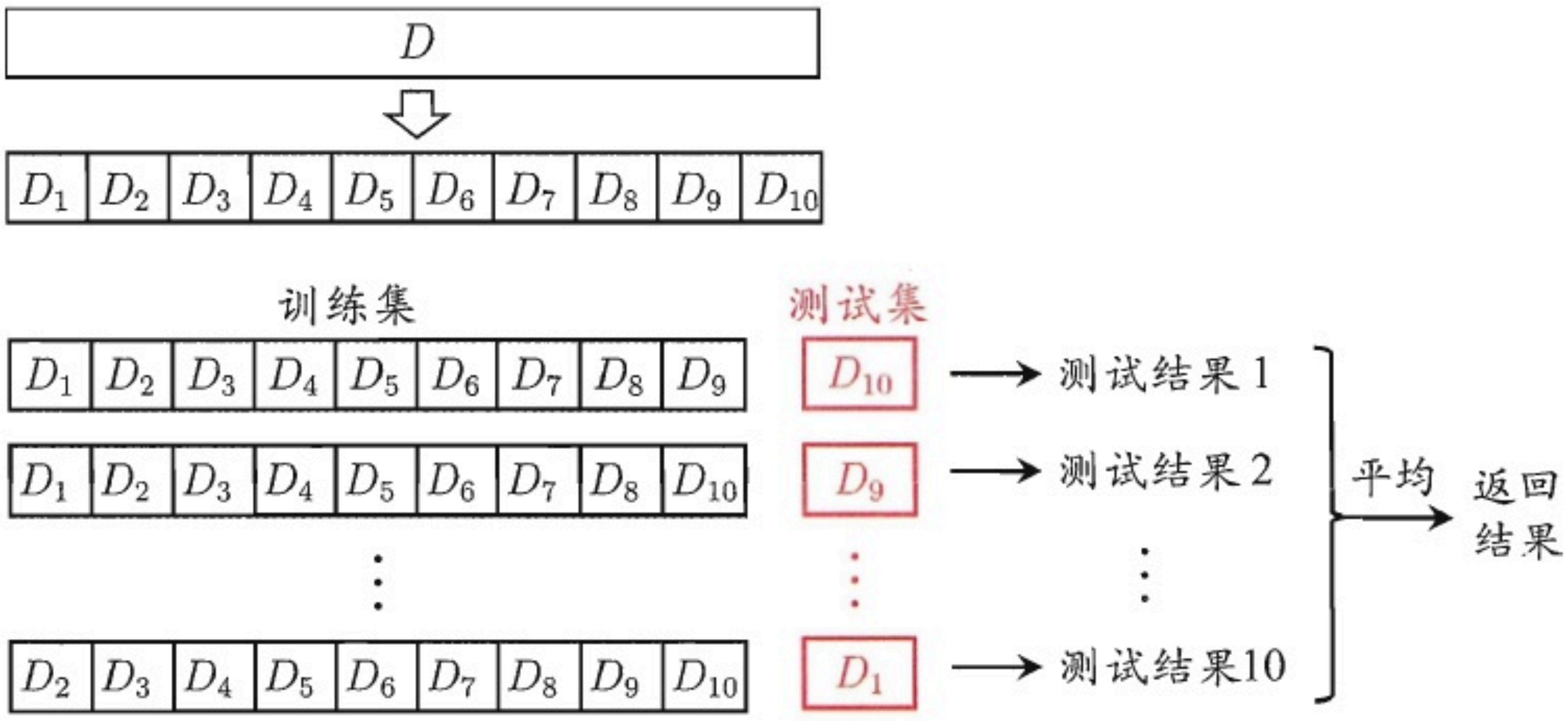
R^2 = sample R-square

p = Number of predictors

N = Total sample size.

- Mellow Cp
- AIC、BIC
- Validation selection
- N-fold validation selection 交叉验证

K折交叉**验证**


D
D_1 D_2 D_3 D_4 D_5 D_6 D_7 D_8 D_9 D_{10}
训练集
测试集
D_1 D_2 D_3 D_4 D_5 D_6 D_7 D_8 D_9 D_{10} 测试结果 1
D_1 D_2 D_3 D_4 D_5 D_6 D_7 D_8 D_{10} D_9 测试结果 2
D_2 D_3 D_4 D_5 D_6 D_7 D_8 D_9 D_{10} D_1 测试结果10
平均
返回
结果

特征缩减（Shrinkage method）

回归模型： $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$

$$\mathrm{RSS} = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2.$$

Lasso回归： Loss= $\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| = \mathrm{RSS} + \lambda\sum_{j=1}^{p}|\beta_j|.$

当λ ⟶ 无穷， $\hat{\beta} = \begin{pmatrix} \bar{y} \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$

特征缩减（Shrinkage method）

回归模型： $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$

$$\mathrm{RSS} = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 .$$

岭回归： Loss= $\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 = \mathrm{RSS} + \lambda \sum_{j=1}^{p} \beta_j^2,$

$$L(w)=\sum_{i=1}^{N}\|w^Tx_i-y_i\|^2=\sum_{i=1}^{N}(w^Tx_i-y_i)^2$$

$$=\underbrace{\begin{pmatrix}w^Tx_1-y_1 & w^Tx_2-y_2 & \cdots & w^Tx_N-y_N\end{pmatrix}}_{\shortparallel}\underbrace{\begin{pmatrix}w^Tx_1-y_1\\ w^Tx_2-y_2\\ \vdots\\ w^Tx_N-y_N\end{pmatrix}}_{=XW-Y}$$

$$(w^Tx_1\ w^Tx_2\ \cdots\ w^Tx_N)-(y_1\ y_2\ \cdots\ y_N)$$

$$=w^T(x_1\ x_2\ \cdots\ x_N)-(y_1\ y_2\ \cdots\ y_N)$$

$$=W^TX^T-Y^T$$

矩阵导数！

$X\in\mathbb{R}^{N\times p}$

$Y\in\mathbb{R}^{N\times 1}$

$W\in\mathbb{R}^{p\times 1}$

$$L(w)=(W^TX^T-Y^T)(XW-Y)$$

$$=W^TX^TXW-Y^TXW-\underbrace{W^TX^TY}_{\in\mathbb{R}}+Y^TY$$

$$=W^TX^TXW-2W^TX^TY+Y^TY$$

$$\hat{w}=\arg\min L(w)$$

$$\frac{\partial L(w)}{\partial W}=2X^TXW-2X^TY\triangleq 0\Rightarrow X^TXW=X^TY\Rightarrow W=\underbrace{(X^TX)^{-1}X^T}_{X^+\text{(伪逆)}}Y$$

带正则化的线性回归损失函数

$$J(w)=\sum_{i=1}^{N}\|w^Tx_i-y_i\|^2+\lambda w^Tw$$

$$=(W^TX^T-Y^T)(XW-Y)+\lambda w^TW$$

$$=W^TX^TXW-2W^TX^TY+Y^TY+\lambda W^TW$$

$$=W^T(X^TX+\lambda I)W-2W^TX^TY+Y^TY$$

$$\hat{W}=\underset{W}{\arg\min}\,J(w)$$

$$\frac{\partial J(w)}{\partial W}=2(X^TX+\lambda I)W-2X^TY=0$$

$$\hat{W}=(X^TX+\lambda I)^{-1}X^TY.$$

岭回归下的$\hat{W}$：$\hat{W}=(X^TX+\lambda I)^{-1}X^TY.$

线性回归的$\hat{W}$：$\hat{W}=(X^TX)^{-1}X^TY$

特征降维（PCR）

PCA：$X_1, X_2, \dots, X_p \rightarrow Z_1, Z_2, \dots, Z_m$

找到p*m维的Φ，使得$X * \Phi = Z$

$$\begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \phi_{11} & \cdots & \phi_{1m} \\ \vdots & \ddots & \vdots \\ \phi_{\mathrm{p}1} & \cdots & \phi_{pm} \end{bmatrix} = \begin{bmatrix} \mathrm{z}_{11} & \cdots & z_{1m} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nm} \end{bmatrix}$$

如何求Φ？

PCA：$X_1, X_2, \dots, X_p \ \rightarrow \ Z_1, Z_2, \dots, Z_m$

找到p*m维的ϕ，使得$X * \Phi = Z$

$$\begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \phi_{11} & \cdots & \phi_{1m} \\ \vdots & \ddots & \vdots \\ \phi_{\mathrm{p}1} & \cdots & \phi_{pm} \end{bmatrix} = \begin{bmatrix} \mathrm{z}_{11} & \cdots & z_{1m} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nm} \end{bmatrix}$$

如何求Φ？

$$\Phi_1 = [\begin{matrix} \phi_{11} \\ \vdots \\ \phi_{p1} \end{matrix}] \qquad \begin{cases} Z_1 = X * \Phi_1,\ \max(Var(Z_1)) \\ {\Phi_1}^T * \Phi_1 = 1 \end{cases}$$

$$\Phi_2 = [\begin{matrix} \phi_{12} \\ \vdots \\ \phi_{p2} \end{matrix}] \qquad \begin{cases} Z_2 = X * \Phi_2,\ \max(Var(Z_2)) \\ {\Phi_2}^T * \Phi_2 = 1 \\ \Phi_1\text{和}\Phi_2\text{不相关，}{\Phi_2}^T * \Phi_1 = 0 \end{cases}$$

$$\vdots \qquad\qquad \vdots$$

$$\Phi_{\mathrm{m}} = [\begin{matrix} \phi_{1m} \\ \vdots \\ \phi_{pm} \end{matrix}] \qquad \begin{cases} Z_m = X * \Phi_m,\ \max(Var(Z_m)) \\ {\Phi_m}^T * \Phi_m = 1 \\ \Phi_1, \Phi_2, \Phi_3 \dots, \Phi_m\text{不相关} \end{cases}$$

PCA：$X_1, X_2, \ldots, X_p \rightarrow Z_1, Z_2, \ldots, Z_m$

找到p*m维的ϕ，使得$X * \Phi = Z$

$$\begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \phi_{11} & \cdots & \phi_{1m} \\ \vdots & \ddots & \vdots \\ \phi_{\mathrm{p}1} & \cdots & \phi_{pm} \end{bmatrix} = \begin{bmatrix} \mathrm{z}_{11} & \cdots & z_{1m} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nm} \end{bmatrix}$$

PCR：$X_1, X_2, \ldots, X_p \rightarrow Z_1, Z_2, \ldots, Z_m \rightarrow$ 拿 [$Z_1, Z_2, \ldots, Z_m$] 和 Y变量建立回归模型。

特征降维（PLS）

$$\Phi_1 = \begin{bmatrix} \phi_{11} \\ \vdots \\ \phi_{p1} \end{bmatrix} \qquad \begin{cases} Z_1 = X * \Phi_1,\ \max(\rho_{Z_1,Y}) \\ {\Phi_1}^T * \Phi_1 = 1 \end{cases}$$

$$\Phi_2 = \begin{bmatrix} \phi_{12} \\ \vdots \\ \phi_{p2} \end{bmatrix} \qquad \begin{cases} Z_2 = X * \Phi_2,\ \max(\rho_{Z_2,Y}) \\ {\Phi_2}^T * \Phi_2 = 1 \\ \Phi_1\text{和}\Phi_2\text{不相关，}{\Phi_2}^T * \Phi_1 = 0 \end{cases}$$

$$\vdots \qquad \vdots$$

$$\Phi_{\mathrm{m}} = \begin{bmatrix} \phi_{1m} \\ \vdots \\ \phi_{pm} \end{bmatrix} \qquad \begin{cases} Z_m = X * \Phi_m,\ \max(\rho_{Z_m,Y}) \\ {\Phi_m}^T * \Phi_m = 1 \\ \Phi_1, \Phi_2, \Phi_3 \ldots, \Phi_m\text{不相关} \end{cases}$$

PLS： 通过 $\max(\rho_{Z_\mathrm{i},Y})$,$where\ Z_i = X * \Phi_i$ ， 建立回归模型。

$$\Phi_1 = \begin{bmatrix} \phi_{11} \\ \vdots \\ \phi_{p1} \end{bmatrix} \qquad \begin{cases} Z_1 = X * \Phi_1，\ \max(\rho_{Z_1,Y}) \\ {\Phi_1}^T * \Phi_1 = 1 \end{cases}$$

$$\Phi_2 = \begin{bmatrix} \phi_{12} \\ \vdots \\ \phi_{p2} \end{bmatrix} \qquad \begin{cases} Z_2 = X * \Phi_2，\ \max(\rho_{Z_2,Y}) \\ {\Phi_2}^T * \Phi_2 = 1 \\ \Phi_1\text{和}\Phi_2\text{不相关，}{\Phi_2}^T * \Phi_1 = 0 \end{cases}$$

$$\vdots \qquad\qquad \vdots$$

$$\Phi_\mathrm{m} = \begin{bmatrix} \phi_{1m} \\ \vdots \\ \phi_{pm} \end{bmatrix} \qquad \begin{cases} Z_m = X * \Phi_m，\ \max(\rho_{Z_m,Y}) \\ {\Phi_m}^T * \Phi_m = 1 \\ \Phi_1, \Phi_2, \Phi_3 \dots, \Phi_m\text{不相关} \end{cases}$$