因为特征排列顺序并无特殊限制,不妨取 x_p 为因变量,其它特征为自变量,建立多元线性回归模型

$$oldsymbol{x}_p = lpha_1 oldsymbol{x}_1 + lpha_2 oldsymbol{x}_2 + \cdots + lpha_{p-1} oldsymbol{x}_{p-1} + oldsymbol{\epsilon}$$

令 $\boldsymbol{X}_t = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{p-1}), \ \boldsymbol{X}_t$ 是该回归模型的构造矩阵

$$\hat{oldsymbol{x}}_p = oldsymbol{X}_t \hat{oldsymbol{eta}}_t = oldsymbol{X}_t (oldsymbol{X}_t' oldsymbol{X}_t)^{-1} oldsymbol{X}_t' oldsymbol{x}_p$$

记 $\boldsymbol{H} = \boldsymbol{X}_t (\boldsymbol{X}_t' \boldsymbol{X}_t)^{-1} \boldsymbol{X}_t', \ \boldsymbol{H}$ 是一个帽子矩阵

回归系数
$$R_p^2 = rac{SS_{Rp}}{SS_{Tp}}$$

其中

$$egin{aligned} SS_{Rp} &= \sum_{i=1}^n (\hat{x}_{ip} - \overline{oldsymbol{x}}_p)^2 = \sum_{i=1}^n \hat{x}_{ip}^2 = \hat{oldsymbol{x}}_p' \hat{oldsymbol{x}}_p = (oldsymbol{H} oldsymbol{x}_p)'(oldsymbol{H} oldsymbol{x}_p) = oldsymbol{x}_p' oldsymbol{H}' oldsymbol{H} oldsymbol{x}_p = oldsymbol{x}_p' oldsymbol{H} oldsymbol{x}_p \ SS_{Tp} = \sum_{i=1}^n (x_{ip} - \overline{oldsymbol{x}}_p)^2 = \sum_{i=1}^n x_{ip}^2 = 1 \ &\Rightarrow R_p^2 = oldsymbol{x}_p' oldsymbol{H} oldsymbol{x}_p \end{aligned}$$

因为 $oldsymbol{X}_s = (oldsymbol{X}_t, oldsymbol{x}_p)$

$$oldsymbol{X}_s'oldsymbol{X}_s = egin{pmatrix} oldsymbol{X}_t' \ oldsymbol{x}_p' \end{pmatrix} (oldsymbol{X}_t & oldsymbol{x}_p) = egin{pmatrix} oldsymbol{X}_t'oldsymbol{X}_t & oldsymbol{X}_t'oldsymbol{x}_p \ oldsymbol{x}_p'oldsymbol{X}_t & oldsymbol{x}_p'oldsymbol{x}_p \end{pmatrix}$$

记 $m{X}_t'm{X}_t = m{A}_{11}, \ m{X}_t'm{x}_p = m{A}_{12}, \ m{x}_p'm{X}_t = m{A}_{21}, \ m{x}_p'm{x}_p = m{A}_{22}$

$$\Rightarrow (\boldsymbol{X}_s'\boldsymbol{X}_s)^{-1} = \begin{pmatrix} \boldsymbol{A}_{11}^{-1} + \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{M}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{M}^{-1} \\ -\boldsymbol{M}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{M}^{-1} \end{pmatrix}$$

其中 $\boldsymbol{M} = \boldsymbol{A}_{22} - \boldsymbol{A}_{21} \boldsymbol{A}_{11}^{-1} \boldsymbol{A}_{12}$

因为我们关心的是 VIF_p ,所以只需求解 M^{-1}

$$m{M} = m{A}_{22} - m{A}_{21} m{A}_{11}^{-1} m{A}_{12} = m{x}_p' m{x}_p - m{x}_p' m{X}_t (m{X}_t' m{X}_t)^{-1} m{X}_t' m{x}_p = m{x}_p' m{x}_p - m{x}_p' m{H} m{x}_p = 1 - R_p^2$$
所以 $m{M}^{-1} = (1 - R_p^2)^{-1}$

因为 VIF_p 为 $(\boldsymbol{X}_s'\boldsymbol{X}_s)^{-1}$ 的第p个对角线元素

$$\Rightarrow VIF_p = oldsymbol{M}^{-1} = rac{1}{1-R_p^2}$$

因为特征顺序对模型回归结果无影响,因此将任一特征作为第 p 个特征对结果无影响。

即该结论具有一般性、可推广为

$$VIF_j = rac{1}{1-R_j^2}$$

Q2

由最小二乘的性质可得

$$E(\hat{oldsymbol{eta}}) = oldsymbol{eta}, \quad \mathrm{Var}(\hat{oldsymbol{eta}}) = \sigma^2(oldsymbol{X}'oldsymbol{X})^{-1}$$
 $MSE(\hat{oldsymbol{eta}}) = E(\hat{oldsymbol{eta}} - oldsymbol{eta})'(\hat{oldsymbol{eta}} - oldsymbol{eta})$
 $= E(\hat{oldsymbol{eta}} - E\hat{oldsymbol{eta}})'(\hat{oldsymbol{eta}} - E\hat{oldsymbol{eta}})$
 $= \mathrm{tr}\Big(\mathrm{Cov}(\hat{oldsymbol{eta}})\Big)$
 $= \mathrm{tr}\Big(\mathrm{Var}(\hat{oldsymbol{eta}})\Big)$
 $= \mathrm{tr}\Big(\sigma^2(oldsymbol{X}'oldsymbol{X})^{-1}\Big)$
 $= \sigma^2 \, \mathrm{tr}\, oldsymbol{(X'X)}^{-1}$

设 X'X 的特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_{p+1}$

根据矩阵的性质,有 $(\mathbf{X}'\mathbf{X})^{-1}$ 的特征值为 $\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_{n+1}^{-1}$, 因此

$$\mathrm{tr}\left(oldsymbol{X}'oldsymbol{X}
ight)^{-1} = \sum_{i=1}^{p+1}rac{1}{\lambda_i} \quad \Rightarrow \quad MSE(\hat{oldsymbol{eta}}) = \sigma^2\sum_{i=1}^{p+1}rac{1}{\lambda_i}$$

Q3

赤池信息量准则 AIC 定义为

$$AIC = -2\ln(模型最大似然) + 2(模型独立参数个数)$$

在线性回归模型中,假定数据满足以下分布

$$y \sim \mathcal{N}_n(oldsymbol{X}oldsymbol{eta}, \sigma^2oldsymbol{I})$$

 $y = (y_1, y_2, \dots, y_n)'$ 的联合密度函数为

$$f(oldsymbol{y};oldsymbol{eta},\sigma^2) = rac{1}{(2\pi)^{n/2}|\sigma^2oldsymbol{I}|^{1/2}} \mathrm{exp}igg(-rac{1}{2}(oldsymbol{y}-oldsymbol{X}oldsymbol{eta})'(\sigma^2oldsymbol{I})^{-1}(oldsymbol{y}-oldsymbol{X}oldsymbol{eta})igg)$$

参数 (β, σ^2) 的似然函数为

$$L(oldsymbol{eta}, \sigma^2) = (2\pi)^{-n/2}(\sigma^2)^{-n/2} \expigg(-rac{1}{2\sigma^2}(oldsymbol{y} - oldsymbol{X}oldsymbol{eta})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta})igg)$$

对数似然函数为

$$\ln L(oldsymbol{eta},\sigma^2) = -rac{n}{2} \mathrm{ln}(2\pi) - rac{n}{2} \mathrm{ln}ig(\sigma^2ig) - rac{1}{2\sigma^2(oldsymbol{y} - oldsymbol{Xeta})'(oldsymbol{y} - oldsymbol{Xeta})'}$$

β 的最大似然估计

$$egin{aligned} rac{\partial \ln L(oldsymbol{eta}, \sigma^2)}{\partial oldsymbol{eta}} &= rac{\partial (rac{1}{2\sigma^2} (oldsymbol{y} - oldsymbol{X}oldsymbol{eta})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta})}{\partial oldsymbol{eta}} &= rac{1}{\sigma^2} (-oldsymbol{X})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta}) = 0 \ & \hat{oldsymbol{eta}}_{
m ML} &= (oldsymbol{X}'oldsymbol{X})^{-1}oldsymbol{X}'oldsymbol{y} \end{aligned}$$

• σ^2 的最大似然估计

$$egin{aligned} rac{\partial \ln L(oldsymbol{eta}, \sigma^2)}{\partial \sigma^2} &= -rac{n}{2\sigma^2} + rac{1}{2\sigma^4} (oldsymbol{y} - oldsymbol{X}oldsymbol{eta})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta}) &= 0 \ \hat{\sigma}_{ ext{ML}}^2 &= rac{1}{n} (oldsymbol{y} - oldsymbol{X}\hat{oldsymbol{eta}}_{ ext{ML}})'(oldsymbol{y} - oldsymbol{X}\hat{oldsymbol{eta}}_{ ext{ML}}) &= rac{1}{n} (oldsymbol{y} - \hat{oldsymbol{y}})'(oldsymbol{y} - oldsymbol{\hat{y}}) &= rac{1}{n} oldsymbol{e}'oldsymbol{e} \end{aligned}$$

将参数估计代入,得到最大对数似然函数

$$\ln L(oldsymbol{eta},\sigma^2) = -rac{n}{2} \mathrm{ln}(2\pi) - rac{n}{2} \mathrm{ln} \, rac{oldsymbol{e}'oldsymbol{e}}{n} - rac{n}{2oldsymbol{e}'oldsymbol{e}} (oldsymbol{y} - oldsymbol{X}\hat{oldsymbol{eta}})'(oldsymbol{y} - oldsymbol{X}\hat{oldsymbol{eta}})'$$

因此

$$egin{aligned} ext{AIC} &= -2 \ln L(oldsymbol{eta}, \sigma^2) + 2(p+2) \ &= n \ln(2\pi) + n \ln rac{oldsymbol{e}' oldsymbol{e}}{n} + rac{n}{oldsymbol{e}' oldsymbol{e}} (oldsymbol{y} - oldsymbol{X} \hat{oldsymbol{eta}})'(oldsymbol{y} - oldsymbol{X} \hat{oldsymbol{eta}}) + 2(p+2) \ &= n \ln(2\pi) + n \ln \left(rac{oldsymbol{e}' oldsymbol{e}}{n} + rac{n}{oldsymbol{e}' oldsymbol{e}} (oldsymbol{y} - \hat{oldsymbol{y}})'(oldsymbol{y} - oldsymbol{\hat{y}})' + 2(p+2) \ &= n \ln(2\pi) + n \ln \left(rac{SS_E}{n}
ight) + n + 2(p+2) \end{aligned}$$

04

已知岭回归估计是最小化带有 L_2 正则项的离差平方和的解,即

$$\hat{oldsymbol{eta}}(k) = rg \min_{oldsymbol{eta}} (oldsymbol{y} - oldsymbol{X}oldsymbol{eta})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta}) + \lambda oldsymbol{eta}'oldsymbol{eta} \qquad \cdots \cdots (1)$$

以下通过贝叶斯统计中的最大后验估计证明上式

首先考虑线性回归模型

$$oldsymbol{y} = oldsymbol{X}oldsymbol{eta} + oldsymbol{\epsilon}$$

其中,y是中心化后的响应变量,X是标准化后的特征矩阵, β 是待估参数,误差项 $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ 在这个模型下,给定参数 β ,响应变量 y 的似然函数为

$$egin{aligned} P(oldsymbol{y}|oldsymbol{eta}) &= rac{1}{(2\pi)^{n/2}|\sigma^2oldsymbol{I}|^{1/2}} \mathrm{exp}igg(-rac{1}{2}(oldsymbol{y}-oldsymbol{X}oldsymbol{eta})'(\sigma^2oldsymbol{I})^{-1}(oldsymbol{y}-oldsymbol{X}oldsymbol{eta}) \ &= (2\pi)^{-n/2}(\sigma^2)^{-n/2} \, \mathrm{exp}igg(-rac{1}{2\sigma^2}(oldsymbol{y}-oldsymbol{X}oldsymbol{eta})'(oldsymbol{y}-oldsymbol{X}oldsymbol{eta}) \ \end{aligned}$$

对于 β_i ,我们选择正态分布作为其先验分布

$$eta_i \sim \mathcal{N}(0, au^2), \quad i=1,\cdots,p$$

即 $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{0}, \tau^2 \boldsymbol{I})$, 所以有

$$P(oldsymbol{eta}) = (2\pi)^{-p/2} (au^2)^{-p/2} \expigg(-rac{1}{2 au^2}oldsymbol{eta}'oldsymbol{eta}igg)$$

根据贝叶斯定理, 后验分布为

$$egin{aligned} P(oldsymbol{eta}|oldsymbol{y}) &= P(oldsymbol{y}|oldsymbol{eta}) \cdot P(oldsymbol{eta}) \ &= (2\pi)^{-n/2}(\sigma^2)^{-n/2} \expigg(-rac{1}{2\sigma^2}(oldsymbol{y} - oldsymbol{X}oldsymbol{eta})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta}) \\ &\propto \expigg(-rac{(oldsymbol{y} - oldsymbol{X}oldsymbol{eta})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta})}{2\sigma^2} - rac{oldsymbol{eta}'oldsymbol{eta}}{2\tau^2}igg) \end{aligned}$$

要最大化后验分布, 考虑最大化其对数函数

$$L(oldsymbol{eta}) = -rac{1}{2}igg(rac{(oldsymbol{y}-oldsymbol{X}oldsymbol{eta})'(oldsymbol{y}-oldsymbol{X}oldsymbol{eta})}{\sigma^2} + rac{oldsymbol{eta}'oldsymbol{eta}}{ au^2}igg)$$

可转化为最小化 $-2L(\beta)$

$$ilde{L}(oldsymbol{eta}) = -2L(oldsymbol{eta}) = rac{1}{\sigma^2}(oldsymbol{y} - oldsymbol{X}oldsymbol{eta})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta}) + rac{1}{ au^2}oldsymbol{eta}'oldsymbol{eta}$$

令 $\frac{\sigma^2}{\tau^2} = \lambda$,所以得到后验分布的最大值

$$rg \min_{oldsymbol{eta}} ilde{L}(oldsymbol{eta}) = rg \min_{oldsymbol{eta}} \left(rac{1}{\sigma^2} (oldsymbol{y} - oldsymbol{X}oldsymbol{eta})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta})'(oldsymbol{y} - oldsymbol{X}oldsymbol{eta}) + \lambdaoldsymbol{eta}'oldsymbol{eta})
ight. \cdots \cdots (2)$$

比较 (1), (2) 两式,形式相同,因此从贝叶斯统计的角度,岭回归可以看作是在进行线性回归时,通过对模型参数施加一个正态先验分布,从而引入了 L_2 正则化。