SUR and classification

Chengwei Li
Sudong Chen

catalogue

1.An introduction to SUR

2.SUR and multiclass classification tasks

1. An introduction to SUR

SUR

SUR(**seemingly unrelated regressions**)

concept：

- regression： one of linear regression models
- **seemingly unrelated:**
  - there is correlation between the different obsevations, y_i, in the dataset
  - **metrics:** $cov(y_i, y_j) \neq 0$, the covariance of the y_i between different equations at the same stage is not equal to zero

advantages:

- Comparing to ignore the covariance, take the covariance into consideration can get the BLUE(best linear unbiases estimator)

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{pmatrix} = X\beta + e$$

m is the number of objects to be predicted,

each y / x / β / e is a sequence of T stages(/T sample points)

SUR

suppose m= 2

$$\sum = V(e) = \begin{pmatrix} \sigma_{11} I_T & \sigma_{12} I_T \\ \sigma_{21} I_T & \sigma_{22} I_T \end{pmatrix}$$

$(\sigma_{ij} = \mathrm{cov}(y_i, y_j))$

if $\sum$ is not a diagonal matrix:

$$\Rightarrow \tilde{y} = \Sigma^{-1/2} y, \tilde{X} = \Sigma^{-1/2} X, \tilde{e} = \Sigma^{-1/2} e$$

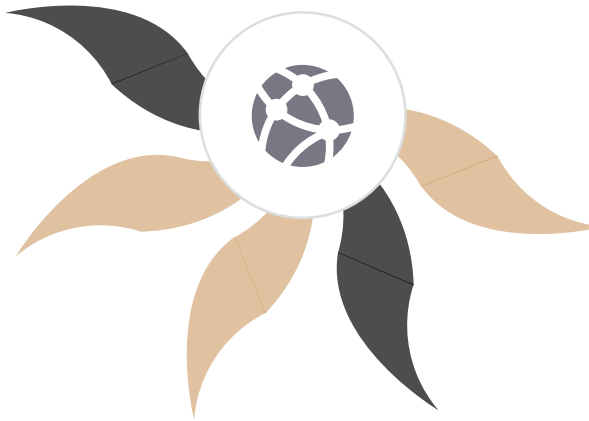
least square method

objective function: $\tilde{f} = \| \tilde{e} \|^2 = \| \tilde{y} - \tilde{X} \beta \|^2 = (y - X\beta)^\top \Sigma^{-1} (y - X\beta)$

$$\arg\min_{\beta} \tilde{f}$$

$$\Rightarrow \frac{d \tilde{f}}{d \beta} = 0$$

$$\Rightarrow \beta^* = (X^\top \Sigma^{-1} X)^{-1} X^\top \Sigma^{-1} y$$



2. SUR and multiclass classification tasks

Current work

$$\beta^* = (X^\top \Sigma^{-1} X)^{-1} X^\top \Sigma^{-1} y$$

assume **scalar is a special case of vector, thus linear regression equation can be used to solve multiclass classification tasks**

1. encode the classes

$\Rightarrow$ *one* hot encoding

(*e.g. class one* $(0,0,1)$, *class two* $(0,1,0)$, *class three* $(1,0,0)$)

2. train multiple classifiers simultaneously

(one equation of SUR is on behalf of one classifier)

3. merge the results to one

Current work

$$\beta^* = (X^\top \Sigma^{-1} X)^{-1} X^\top \Sigma^{-1} y$$

dimention change:

$y : 1 \times 1 \rightarrow 1 \times c$

(c is the number of classes)

$\Rightarrow \sum : mT \times mT \rightarrow mT\mathrm{c} \times mT\mathrm{c}$

(can't pre-multiply by $X^\top (mT \times \sum_u^m l_u)$)

Q: how to converse the $\mathrm{c} \times \mathrm{c}$ matrix to 1×1

End