HMMA 307 : Advanced Linear Modeling

Chapter 1 : Linear regression

Santinelli Emma Mégane Dieval Yassine Sayd

https://github.com/MegDie/advanced_lm_introduction

Université de Montpellier

Table of Contents

Table of Contents

Model

Suppose the data consists of n observations $(yi, xi)_{i=1}^{n}$ with p features. The model can be written in matrix notation as :

$$y = X\beta + \epsilon$$

where

- X is an $n \times p$ matrix of regressors
- β is a p$\times$1 vector of unknown parameters
- ϵ is a vector of normal random errors with mean 0

The OLS estimator is any coefficient vector $\hat{\beta}^{LS} \in \mathbb{R}^p$ such that :

$$\hat{\beta}^{LS} \in argmin \frac{1}{2n} \|y - X\beta\|^2$$

with,

$$f(\beta) = \frac{1}{2n} \sum_{i=1}^{n} (y_i - \frac{1}{2n}(X\beta)_i)^2$$

$$= \beta^T \frac{X^T X}{2n} \beta + \frac{1}{2n} \|y\|^2 - \langle y, X\beta \rangle$$

where, $\langle y, X\beta \rangle = y^T X\beta = \beta^T X^T y = \langle \beta, X^T y \rangle$

Notation

The matrix $\hat{\Sigma} = \frac{X^T X}{n}$ matrix is called the Gram matrix.

$$X^T X = \begin{pmatrix} x_1^T \\ . \\ . \\ x_p^T \end{pmatrix} (x_1^T ... x_p^T),$$

The Gram matrix is equivalent to :

$$[X^T X]_{j,j'} = [\langle x_j, x_{j'} \rangle]_{(j,j') \in [1,p]^2}$$

Remark

Most of the times, we scale features.

We have : $\bar{X}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ (1)

To center explanatory variables, we use the equation (1) to build the centered vector X_c

$X_c \Leftarrow X$ - $(\bar{X}_1 1_n, ..., \bar{X}_p 1_n)$ where $1_n = (1, ..., 1)$

Then we obtain $\bar{X}_c = O_n$

To reduce explanatory variables, we use :

$$\hat{\sigma}_j^2 = \frac{1}{n}\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)$$

Let Xr be the reduced vector, then :

$$Xr_j = \frac{Xj - \bar{X}_j 1_n}{\hat{\sigma}_j}$$

First Order Optimality Condition

We can verify the first order optimality condition because $\nabla f(\hat{\beta}^{LS}) = 0$
Note that f is a C^∞ function, then differentiable

Remark

f is a convex function, so f has a local minimum and a global one.

Conclusion

$\hat{\beta}^{LS}$ satisfy the following equations of orthogonality :

- $\frac{X^T X}{n}\hat{\beta}^{LS} - \frac{X^T y}{n} = 0$
- $\iff \mathrm{X}^T(\frac{X\hat{\beta}^{LS} - y}{n}) = 0$
- $\iff \mathrm{X}^T(y - X\hat{\beta}^{LS}) = 0$
- $\iff \langle X_j, y - X\beta \rangle = 0$ for j in 1:p

Attention

If $p < n$ so rank(X) $\leq n <$ p Then $\hat{\beta}^{LS}$ is not unique

interpretation

- Each explanatory variable is orthogonal to the residuals $\Gamma = y - X\hat{\beta}^{LS}$ With $\hat{\beta}^{LS}$ is a solution of the linear *pxp* system :

$$\hat{\Sigma}\beta = \frac{X^T y}{n}$$

Remark

- If $\hat{\Sigma}$ is invertible, the solution of the linear system is unique
- $\hat{\Sigma}$ is invertible $\Leftrightarrow$ $\hat{\Sigma}$ is positive definite
- if $\hat{\Sigma}$ invertible, so $rank(\hat{\Sigma}) = p$
- we assume thatwe have a full rank column e.g. :

$$rg(X) = dim(Vect(X_1, ..., X_p)) \leq n$$

Remark

- If $rank(X) = p$, so $\hat{\Sigma}$ is invertible and :

$$\hat{\beta}^{LS} = \hat{\Sigma}^{-1}\frac{X^T y}{n} = (\frac{X^T y}{n})^{-1}\frac{X^T y}{n}$$

so :

$$\hat{\beta}^{LS} = (X^T X)^{-1} X^T y$$

Notice

- In practice it is exceptional to invert $\hat{\Sigma}$ because one solves many linear systems

Table of Contents

Reminder

Let $\Sigma \in \mathbb{R}^{p*p}$.
If $\Sigma^T = \Sigma$ then Σ is diagonalizable.

Theorem

For all matrix $M \in \mathbb{R}^{m1*m2}$ of rank r, there exist two orthogonal matrix $U \in \mathbb{R}^{m1*r}$ and $V \in \mathbb{R}^{m2*r}$ such that :

$$M = Udiag(s_1 ... s_r)U^T$$

where $s_1 \geq s_2 \geq ... \geq s_r \geq 0$ are the singular values of M.

Note that : $M = \sum_{j=1}^{r} s_j u_j v_j^T$ with : $U = [u_1, ..., u_r]$ et $V = [v_1 ... v_r]$

Definition

For $M \in \mathbb{R}^{m1*m2}$, a pseudoinverse of M is defined as a matrix M^+ satisfying :

$$M^+ = Vdiag(\frac{1}{s_1}...\frac{1}{s_r})U^T = \sum_{j=1}^{r} \frac{1}{s_j} v_j u_j^T$$

Remark : If M is invertible, its pseudoinverse is its inverse. That is, $A^+ = A^{-1}$

Bibliography

[1] Joseph Salmon, *Modéle linéaire avancé : introduction*, 2019,
http://josephsalmon.eu/enseignement/Montpellier/HMMA307/Introduction.pdf.

[2] Francois Portier and Anne Sabourin, *Lecture notes on ordinary least squares*, 2019,
https://perso.telecom-paristech.fr/sabourin/mdi720/main.pdf

[3] *Ordinary least squares*, 2020,
https://en.wikipedia.org/wiki/Ordinary_least_squares.

[4] *Singular value decomposition*, 2020,
https://en.wikipedia.org/wiki/Singular_value_decomposition.