Week 1 : Advanced Statistics & Regression

Course 2 : Multivariate Linear Regression

Reminders :

- Random Vector :
  X is random vector if each coordinate X_i is random variable
  So : $\mathbb{E}(X)$ = ($\mathbb{E}(X_1)$, ..., $\mathbb{E}(X_d)$)

- Affine transformation :
  if : $Y = \alpha + AX$ so : $\mathbb{E}(Y)$ = $\alpha + A\,\mathbb{E}(X)$
  and $Cov(Y) = ACov(X)\mathbf{A}^\top$

- Gaussian Vector (GV) :
  X is a GV if, and only if, its characteristic function is :
  $\Phi(t) = exp(im - \frac{t^2\sigma^2}{2})$
  Or, for all $a \in R^d$, we have $\mathbf{a}^\top X$ is a guassian variable.

- Standard Gaussian Vector :
  X is standrad GV if $\mathbb{E}(X) = 0$ and $Cov(X) = I_d$

- if X is GV so all X_i are guassian variable and X_i is independent of X_j if $Cov(X_i, X_j) = 0$

- Cochran's Theorem : c.f slides

Multivariate Regression :

Goal :
create a statistical framework for :

- Probabilistic model that describe the observations
- Construction of estimators
- Statistic optimality
- Statistic inference

Linear Model : $Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} + \epsilon_i$ with $i = 1, ..., n$

Simply : $Y = X\beta + \epsilon$

P.s : The model is lineary dependent to β not necessarily to X

- Model is identified if, and only if, $X\mathbf{X}^\top$ is invertible (X is of rank full).

Hypo

- $\mathbb{E}(\epsilon_i) = 0$: errors are centralized
- $Var(\epsilon_i) = \sigma^2$: errors with constant variance
- $Cov(\epsilon_i, \epsilon_j) = 0$: The errors are decorrelated from each other

But we can simply suppose that $\epsilon_i \sim \text{Norm}(0, \sigma^2)$

- **LSE**
  $\hat{\beta} = argmin_{u \in R^p} \|Y - Xu\|^2$
  with norm is the euclidian norm.

So, if $X\mathbf{X}^\top$ is invertible : $\hat{\beta} = (X\mathbf{X}^\top)^{-1}\mathbf{X}^\top Y$

- Gauss-Markov's Theorem :

if $X\mathbf{X}^\top$ is invertible, the estimor $\hat{\beta}$ LSE is guaranting the minimal variance within all the linear estimors.

Plus, It's unbiased : $\mathbb{E}_\beta(\hat{\beta}) = \beta$
and $Var_\beta(\hat{\beta}) = \sigma^2(X\mathbf{X}^\top)^{-1}$

- We define empirical residuals as : $\hat{\epsilon} = Y - X\hat{\beta}$
  so the estimation of σ wih LSE is given by : $\hat{\sigma}^2 = \frac{\|\hat{\epsilon}\|^2}{n-p}$
- For guassian model : LSE $\sim$ MLE. However, MLE is biaised given $\hat{\sigma}^2_{ML} = \frac{\|\hat{\epsilon}\|^2}{n}$

Hypothesis Tests :

for this part, we will present only the results, and the slides give wider explaination yet no proof is presented.

- Distribution of LSE :
  If all hypothesis above are met, with $X\mathbf{X}^\top$ is invertible,
  we have, for all $c \in R^p$: $\frac{\mathbf{c}^\top\hat{\beta} - \mathbf{c}^\top\beta}{\hat{\sigma}\sqrt{\mathbf{c}^\top(X\mathbf{X}^\top)^{-1}c}} \sim t_{(n-p)}$

with $t_{(n-p)}$ is the student law with n-p degree of freedom

- Trust Region :

from the result above, we can define the trust interval with the level $1 - \alpha$
$I_{c,\alpha} = [\mathbf{c}^\top\hat{\beta} \pm t_{n-p,1-\alpha/2}\hat{\sigma}\sqrt{\mathbf{c}^\top(X\mathbf{X}^\top)^{-1}c}]$

with $t_{n-p,1-\alpha/2}$ is the $1 - \alpha/2$ quantile of $t_{(n-p)}$

- Tests :
  we present here the test presented in the slides, yet the details to be checked there,

so the test admits an error: $\Phi(Y) = 1_{(|T|>t_{n-p,1-\alpha/2})}$

Robust Extensions :

Problem : LSE and MLE do not perform well in the case of $p > n$ (more variables than observations) and in the presence of outliers.

- Huber Loss :
  A solution of outliers, since this loss will be less sensitive to the outliers.
  It's given by :

$$L_\delta(a) = \begin{cases} a^2/2 \text{ if } \mid a \mid < \delta \\ \delta(\mid a \mid - \delta/2) \text{ otherwise.} \end{cases}$$

- Ridge Regression :

The problem becomes : $\hat{\beta} \in argmin_{u \in R^p} ||Y - Xu||^2 + \lambda ||u||^2$

λ : term of regularisation
By tuning λ we control the regularisation
Generally, ridge regression gives better prediction errors
Also, it enhances the numeric stability of the model.