

CorReg

Clément Théry, Christophe Biernacki, Gaétan Loridant

ArcelorMittal Dunkerque, Université de Lille 1, équipe MØdal Inria

8 février 2015

Context

Proposed Models

Structure estimation

Results

Missing values

Tools

1. Steel industry databases.
2. Goal : To understand and prevent quality problems on finished product, knowing the whole process, without a priori.



Regression

$$Y = X\beta + \epsilon \quad (1)$$

OLS

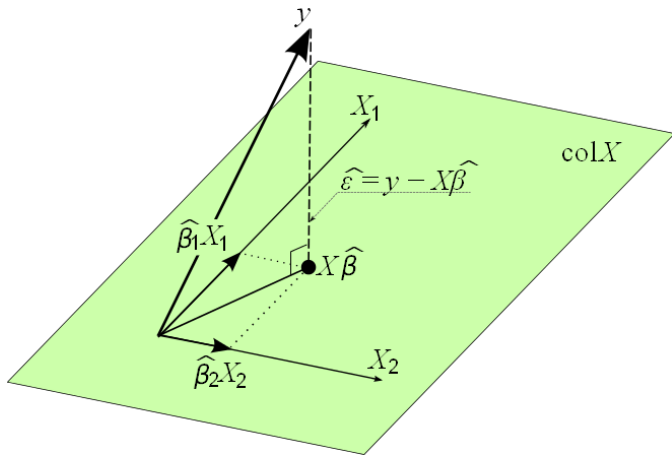


Figure: Multiple linear regression with Ordinary Least Squares seen as a

OLS

β can be estimated by $\hat{\beta}$ with Ordinary Least Squares (ols), that is the unbiased maximum likelihood estimator [Saporta, 2006, Dodge and Rousson, 2004] :

$$\hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad (2)$$

with variance matrix

$$\text{Var}(\hat{\beta}_{OLS}) = \sigma_Y^2 (\mathbf{X}'\mathbf{X})^{-1}. \quad (3)$$

In fact it is the Best Linear Unbiased Estimator (BLUE). The theoretical mse is given by

$$\text{mse}(\hat{\beta}_{OLS}) = \sigma_Y^2 \text{Tr}((\mathbf{X}'\mathbf{X})^{-1}).$$

Ridge

Alternative ridge : bien mais pas de sélection donc inacceptable

Ridge

Alternative ridge : bien mais pas de sélection donc inacceptable

LASSO

Alternative lasso (et autres) : bien mais problèmes en cas de corrélations

LASSO

Liste des alternatives

SEM

Modélisation de la structure mais à la main et aucun impact sur l'estimation

Selvarclust

Semble très bien mais n'aboutit pas vers la régression donc on le prolonge en CorReg













Modèle génératif complet avec dépendances

Explosion des mélanges

SEM avec Gibbs

Bic pondéré

Résultats pourris

Excel, fonctions graphiques, arbres de décision



Dodge, Y. and Rousson, V. (2004).

Analyse de régression appliquée : manuel et exercices corrigés
(coll. eco sup,).

Recherche, 67 :02.



Saporta, G. (2006).

Probabilités, analyse des données et statistique.

Editions Technip.