CorReg

Clément Théry, Christophe Biernacki, Gaétan Loridant

ArcelorMittal Dunkerque, Université de Lille 1,équipe MΘdal Inria

February 8, 2015

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






Regression

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon} \quad (1)$$

where $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma_Y^2 \boldsymbol{I}_n)$

OLS

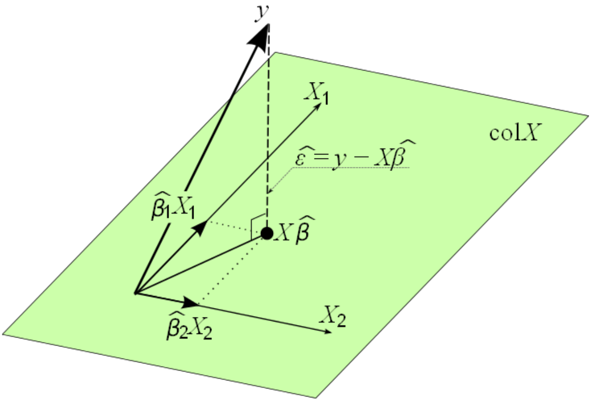

Figure: Multiple linear regression with Ordinary Least Squares seen as a projection on the $d-$dimensional hyperplane spanned by the regressors $\boldsymbol{X}$. Public domain image.

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]:

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

with variance matrix

$$\mathrm{Var}(\hat{\beta}_{OLS}) = \sigma_Y^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{3}$$

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

$$\mathrm{MSE}(\hat{\beta}_{OLS}) = \sigma_Y^2 \,\mathrm{Tr}((\boldsymbol{X}'\boldsymbol{X})^{-1}).$$

Running example

$\boldsymbol{X}^1, \boldsymbol{X}^2, \boldsymbol{X}^4, \boldsymbol{X}^5 \sim \mathcal{N}(0,1)$ and $\boldsymbol{X}^3 = \boldsymbol{X}^1 + \boldsymbol{X}^2 + \varepsilon_1$ where $\varepsilon_1 \sim \mathcal{N}(\mathbf{0}, \sigma_1^2 \boldsymbol{I}_n)$.

Two *scenarii* for $\boldsymbol{Y}$:

$\beta = (1,1,1,1,1)'$ and $\sigma_Y \in \{10, 20\}$.

It is clear that $\boldsymbol{X}'\boldsymbol{X}$ will become more ill-conditioned as σ_1 gets smaller. R^2 stands for the coefficient of determination which is here:

$$R^2 = 1 - \frac{\mathrm{Var}(\varepsilon_1)}{\mathrm{Var}(\boldsymbol{X}^3)} \tag{4}$$

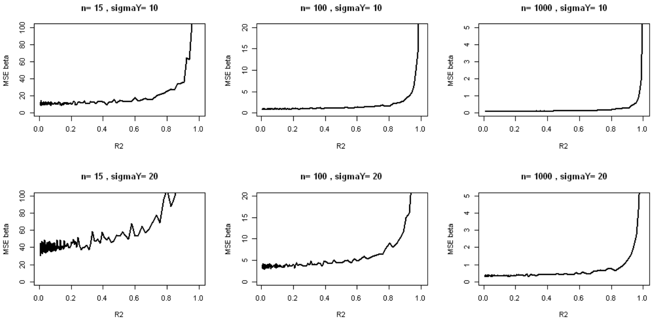

Figure: Evolution of observed Mean Squared error on $\hat{\beta}_{OLS}$ with the strength of the correlations for various sample sizes and strength of regression. $d = 5$ covariates (running example).

Ridge Regression

[Hoerl and Kennard, 1970, Marquardt and Snee, 1975] proposes a possibly biased estimator for β that can be written in terms of a parametric L_2 penalty:

$$\hat{\beta} = \text{argmin}_{\beta} \left\{ \| \boldsymbol{Y} - \boldsymbol{X}\beta \|_2^2 \right\} \text{ subject to } \| \beta \|_2^2 \leq \eta \text{ with } \eta > 0 \tag{5}$$

But this penalty is not guided by correlations. The solution of the ridge regression is given by

$$\hat{\beta} = \left(\boldsymbol{X}'\boldsymbol{X} - \lambda \boldsymbol{I}_n \right)^{-1} \boldsymbol{X}'\boldsymbol{Y} \tag{6}$$

Methods do exist to automatically choose a good value for λ [Cule and De Iorio, 2013, Er et al., 2013] and a R package called `ridge` is on CRAN [Cule, 2014].

Ridge Regression

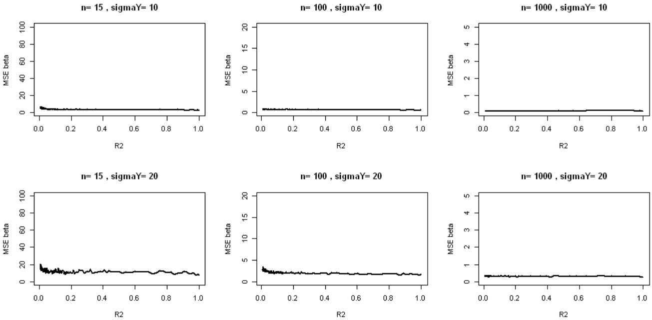

Figure: Evolution of observed Mean Squared error on $\hat{\beta}_{ridge}$ with the strength of the correlations for various sample sizes and strength of regression. $d = 5$ covariates.

LASSO

The Least Absolute Shrinkage and Selection Operator (LASSO, [Tibshirani, 1996] and [Tibshirani et al.,]) consists in a shrinkage of the regression coefficients based on a λ parametric L_1 penalty to obtain zeros in $\hat{\beta}$ instead of the L_2 penalty of the ridge regression:

$$\hat{\boldsymbol{\beta}} = \operatorname{argmin}\left\{\parallel \boldsymbol{Y} - \boldsymbol{X}\beta \parallel_2^2\right\} \text{ subject to } \parallel \boldsymbol{\beta} \parallel_1 \leq \lambda \text{ with } \lambda > 0.$$

LASSO

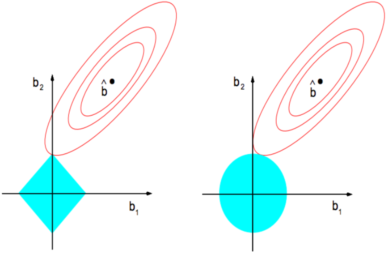

Figure: Geometric view of the Penalty for the LASSO (left) compared to ridge regression (right) as shown in the book from Hastie [Hastie et al., 2009]

Figure shows the contour of error (red) and constraint function (blue). The axis stands for the regression coefficients.

Here again we have to choose a value for λ. The Least Angle Regression (LAR [Efron et al., 2004]) algorithm offers a very efficient way to obtain the whole LASSO path. It can be used through the `lars` package on CRAN ([Hastie and Efron, 2013]).

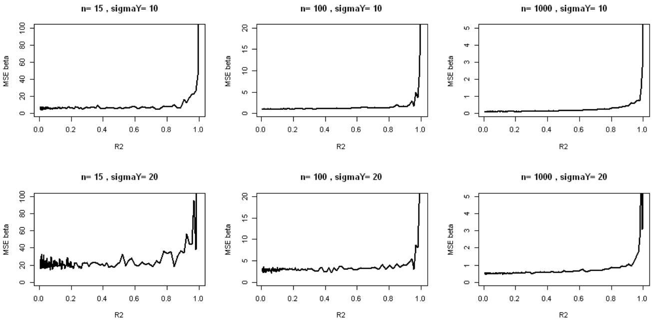

Figure: Evolution of observed Mean Squared error on $\hat{\beta}_{lar}$ with the strength of the correlations for various sample sizes and strength of regression. $d = 5$ covariates.

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

Cule, E. (2014).
ridge: Ridge Regression with automatic selection of the penalty parameter.
R package version 2.1-3.

Cule, E. and De Iorio, M. (2013).
Ridge regression in prediction problems: automatic choice of the ridge parameter.
Genetic epidemiology, 37(7):704–714.

Dodge, Y. and Rousson, V. (2004).
Analyse de régression appliquée: manuel et exercices corrigés (coll. eco sup,).
Recherche, 67:02.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004).
Least angle regression.
The Annals of statistics, 32(2):407–499.