École Nationale de la Statistique et de l'Administration Économique

Rapport du Projet de Statistique

Janvier 2017

Introduction à la régression pénalisée et à l'estimateur lasso

Auteurs :
Mehdi Abbana Bennani
Julien Mattei

Superviseur :
Edwin Grappin

28 décembre 2016

1 Régression OLS

Question 1 : On résout le problème
$min \|\xi\|^2$
On a $\|\xi\|^2 = \|y - X\beta\|^2$

Soit $\nabla_\beta \|y - X\beta\|^2 (\beta^\star) = 0$

D'ou $\nabla_\beta (y - X\beta)^T (y - X\beta)(\beta^\star) = 0$

D'ou $\beta^\star = (X^T X)^{-1} X^T y$
Dans la suite on notera $\beta^\star$ l'estimateur des moindres carrés.
Question 2 :
Voir code.

2 Régression pénalisée : le lasso

Question 3 :
X est une matrice de dimensions n * p
Donc $X^T X$ est une matrice carrée de dimension p
On a $rg(X^T X) = rg(X)$
Donc $rg(X^T X) \leq min(n, p)$
Or vu que le nombre de variables est plus grand que le nombre d'observations, on a $p > n$
Donc $rg(X^T X) < p$
Or $X^T X$ est une matrice de taille p
Donc $X^T X$ n'est pas inversible
Donc on ne peut pas utiliser la méthode précédente.
La matrice $X^T X$ n'est plus inversible car son rang est inférieur au minimum des dimensions de X.
Question 4 :
En dimension 1, on a :
$\|y - X\beta\|^2 = (y - x\beta)^2 = y^2 + \beta^2 x^2 - 2\beta * xy$
Ainsi, le problème revient à trouver β qui minimise la quantité :
$\frac{1}{2}\beta^2 x^2 - \beta * xy + \lambda\beta + y^2$
On dérive par rapport à β :
Si $\beta * x^2 + (\lambda - xy) = 0$
On distingue deux cas dans la résolution :
$\underline{\beta^\star > 0 :}$
En dimension 1 : $\beta^\star = \frac{xy}{x^2}$
Ainsi $\frac{xy - \lambda}{x^2} = \beta^\star - \frac{\lambda}{x^2} = \beta^\star - t$
Avec $t = \frac{\lambda}{x^2}$
Donc : $\hat{\beta}^{lasso} = max(0, \beta^\star - t)$
De même si $\underline{\beta^\star < 0 :}$

$\hat{\beta}^{lasso} = max(0, -\beta^\star - t)$
En résumé : $\hat{\beta}^{lasso} = sign(\beta^\star) * max(0, \beta^\star - t)$
Le seuillage doux permet de regler l'intervalle sur lequel on veut estimer $\beta^\star$. Ainsi, on choisit λ, si notre estimateur $\beta^\star$ est compris dans l'intervalle $[-\lambda, \lambda]$ il est ramené à 0. On ne retient que les valeurs de $\beta^\star$ en dehors de cet intervalle (même principe qu'un filtre).
L'estimateur $\hat{\beta}^{lasso}$ est nul pour tout les $\beta^\star$ tels que $\beta^\star - t < 0$. Donc cela induit une estimation nulle plus souvent que l'estimateur des moindres carrés (qui se produit seulement lorsque $\beta^\star = 0$)

Question 5 : Dans la régression pénalisée, on pénalise les valeurs de β trop grandes, en leur assignant un coût proportionnel à λ PAS REPONDU

Question 6 : Quelles sont les avantages du lasso sur la pénalisation ‘ 0 ? Même question avec la pénalisation Ridge

Question 7 : A implémenter

Question 8 : Le λ permet de régler la "force" de pénalisation du paramètre β.

Question 9 : Dans la méthode de cross validation, on divise le dataset en trois parties : Train Set, Validation Set, et Test Set. On entraine le modèle sur le train set puis on estime la qualité de la prédiction sur le validation set pour plusieurs valeurs de λ, on choisit le lambda le plus performant sur le validation set pour entrainer le modèle sur le nouveau train set comprenant le train set précédent et le validation set.

Question 10 : A implémenter

Combien d'éléments de β estimez vous non-nuls ? Evaluez l'erreur de votre modèle. Trouvez de bonnes mesures pour évaluer votre méthode