

TP 4 : Le lasso composante par composantes (component wise)

Stéphane Canu

2 octobre 2023, APPC, INSA Rouen

Le but du TP est d'étudier différentes solutions au Lasso¹, dans le cadre de la régression sur des données simulées. Pour le faire fonctionner, vous êtes supposé avoir déjà installé CVX (que vous pourrez télécharger à cette adresse : <http://cvxr.com/cvx/>)

Le code suivant est disponible en ligne avec google colab :

<https://colab.research.google.com/drive/1Q1QXbh58oEME1gloZycjhz2FA77HTX5K?usp=sharing>

Ex. 1 — Le Lasso comme une méthode de sélection de variables

Reprendre les données du TP précédent sur le lasso.

1. Coder l'algorithme component wise pour résoudre le Lasso

Pour $j = 1, p$

$$\beta^{(-j)} = (\beta_1, \dots, \beta_{j-1}, 0, \beta_{j+1}, \dots, \beta_p)$$

$$r = y - X\beta^{(-j)}$$

$$\hat{\beta}_j = \arg \min_{\beta \in \mathbb{R}} J^{(j)}(\beta, X_j, r)$$

fin de pour

2. Le Lasso pour sklearn²
 - a) Quels sont les différentes fonction proposée par sklearn pour résoudre le problème du Lasso ? Quelle est l'algorithme implémenté ?
 - b) résoudre le problème à l'aide de la fonction Lasso de sklearn
 - c) résoudre le problème à l'aide de la fonction LassoCV de sklearn
 - d) résoudre le problème à l'aide de la fonction lasso_path de sklearn et visualiser le chemin de régularisation
3. comparez toutes les solutions en terme de temps de calcul et de performances. Proposez et justifiez des recommandations sur la méthode à utiliser pour résoudre un problème de lasso.

¹<http://statweb.stanford.edu/~tibs/lasso.html>

²scikit-learn.org/stable/modules/generated/sklearn.linear_model.Lasso.html