Methode de PCA parcimonieuse

Sofia HARROUCH

30 mars 2018

### SPCA: Sparse Principal Component Analysis

Dans cette partie, on montre une deuxième approche pour l’obtention d’une ACP modifié et “spase”. Cette méthode consiste à faire une analogie entre l’ACP et la regression ridge. Tout d’abord, on montrera que les composantes principales d’une ACP s’écrivent comme un problème d’optimisation d’une regression ridge. Pour ensuite rajouter la pénalité du Lasso afin d’obtenir un problème d’optimisation d’une regression sous la pénalité d’Elastic net ce qui va nous permettre d’avoir des composantes principales parcimonieuse.

#### Approche simple de regression pour l’ACP

On considère la kième composante principale obtenu à partir d’une ACP, qui est une combinaison linéaire des p variables intiales. En effet Ces coefficients de saturation peuvent aussi être obtenu en faisant une regression multiple de la CP sur les p variables initiales (). Ensuite, on étend cette regression à une regression ridge en ajoutant la pénalité Ridge dans le but de manipuler toute sorte de données (quand le nombre de variable est supérieur aux nombres de donnée et la regression multiple n’a pas une solution unique, quand et X n’est pas une matrice complète) et d’obtenir une solution unique comme l’ACP. Pour résumer, cette pénalité ne nous permet juste de reconstruire les composantes principales et non pas de les penaliser. Enfin, on rajoute la pénalité du Lasso pour penaliser les composantes et de les rendre parcimonieuses. Le problème d’optimisation devient: , et la kième composante principale approximée est . Cette technique dépend essentiellement des résulats de l’ACP (car on applique l’ACP et on utilise l’équation (1.1) pour trouver une autre apprixmation approprié des CPs pour qu’elles soient parcimonieuses). On peut dire que cette approche n’est pas une alternative et par la suite pas véritable. Dans la sous partie qui suit, on montrera la même chose que précedement mais cette fois-ci sans l’utilisation du résultat se l’ACP.

#### Approche complexe de regression pour l’ACP

Soient X la matrice de données, et la ième ligne de cette matrice, les k premières composantes principales, et les k premières composantes principales estimées de tel sorte que les CPs soient parcimonieuses. Le théorème 1 suivant montre que le problème de l’ACP peut se transformer en un problème de regression

**Théorème 1** Pour tout :

Sous la contrainte

Alors pour Ce théorème nous permet de bien voir la transformation du pro

En fait,

Comme est orthonormal, est une matrice qualeconque ortonormal tel que $\[A;A\_{ \perp}\]$ est orthonormal. Soit A fixé, le problème de minimisation devient:

Ce qui est équivalent à resoudre k regressions ridges indépendantes. Ce qui donne . En fait, si A correspond aux CPs ordinaire alors on sait effectivement que B est proportionnelle à V (ce qu’on a montré dans la première sous partie).

Aprés avoir montré l’analogie entre les deux méthode, on rajoute la pénalité de lasso pour avoir la “sparsité” des coefficients de regression (de saturation). Le problème devient:

Sous la contrainte:

tel que est utilisé pour les k composantes alors que utilisée pour la “spasité” est différente d’une composante.

#### L’algorithme pour la réslution du problème d’optimisation

* On suppose que A est connu et égale au CPs de l’ACP, et donc:

-On suppose que B est connu et fixé, on ignore la pénalité qui est lié aux termes de , et le problème de minimisation devient:

Soud la contrainte:

La solution est obtenu en utilisant le théorème 2; en calculant la décomposition en valeurs singilieures (SVD) et posant

*Théorème 2* Soit et deux matrices, On considère le problème de minimisation suivant:

Sous la contrainte

Soit la décomposition en valeurs singulière de est donc .

Ceci étant expliqué, l’algorithme peut ^etre résumé comme suit: 1. On commence par les coefficients de saturation des K premiÃ¨res composantes principales. 2. Sachant , On résout le problème pour

1. Pour une matrice fixée, on calcule la décomposition de la matrice en valeur singuliÃ¨re (SVD) de , par la suite on met A à jour, ce qui donne
2. On répète les étapes 2-3 jusqu’à la convergence.
3. On normalise V, Et les sont les nouvelles composantes principales paricimonieuses.