Applications de l’optimisation dans l’apprentissage automatique (machine learning) en économie cas des moindres carrés ordinaires (MCO) et des K-Means.

Axel-Cleris Gailloty

10/11/2020

Applications de l’optimisation dans le champ de l’économie : cas des moindres carrés ordinaires et des K-Moyennes.

# Introduction

L’économie est souvent définie comme la science de la rareté et des choix efficaces. Cette définition part du fait que les besoins des agents économiques sont illimités tandis que les ressources, elles sont limitées et rares. Il revient donc à ce que chaque agent économique définisse sa fonction objectif (fonction d’utilité pour les individus, fonction de production pour les entreprises) à optimiser afin de maximiser son gain compte tenu des moyens qu’il possède.  
Deux branches de la science économique emploient intensivement à l’optimisation : ce sont la microéconomie et l’économétrie.

La microéconomie emploie l’optimisation en vue de construire des théories qui expliquent le comportement humain en face de la rareté, de l’incertitude et dans le cadre des contrats. L’optimisation en microéconomie permet par exemple de trouver le panier optimal de biens à acheter compte tenu du revenu du consommateur. Cette approche permet par exemple de prédire le comportement du consommateur face aux variations de prix. Elle peut aussi aider à déterminer la combinaison optimale des facteurs de production (typiquement le capital et le travail) afin de maximiser le profit ou de réduire les coûts de production.

L’économétrie de son côté utilise tout un ensemble de techniques mathématiques, statistiques et informatiques en vue de tester les théories économiques à travers la modélisation mathématique, l’estimation des paramètres et la prévision.

L’économétrie en tant que branche des sciences économiques a importé dans la science économique un très grand nombre d’outils empruntés à d’autres disciplines comme la biométrie, les statistiques. La méthode des moindres carrés ordinaires qui est aujourd’hui largement utilisé avec des variations mineures dans plusieurs modèles économétriques illustre bien le recours de l’économétrie à d’autres disciplines.

Dans ce travail nous aimerions aborder l’aspect optimisation dans deux algorithmes souvent employés dans des problèmes économiques. Ce sont les moindres carrés ordinaires et les K-Moyennes.  
Nous commencerons par présente chaque algorithme puis nous cherchons deux applications pour chacun.

# Les moindres carrés ordinaires

## Historique de la méthode

La première publication de la méthode des moindres carrés (destinée à déterminer des quantités dans un système d’équations surdéterminé) est due à Legendre qui l’a donnée en annexe d’un ouvrage intitulé *Nouvelles méthodes pour la détermination des orbites des comètes (1805)*

## Formalisation

La méthode des moindres carrés utilise de l’optimisation mathématique. Il s’agit en effet de trouver un vecteur de paramètres qui minimise la **fonction objective** qui est la somme des carrés des résidus.  
Certaines formalisations utilisent une autre fonction objective mais dans la plupart des cas nous cherchons à réduire la somme des carrés des résidus (SCR). L’intérêt de minimiser la SCR est qu’il s’agit d’une fonction convexe qui admet donc un minimum global.  
L’optimisation se fait en utilisant la méthode de Lagrange.

Dans les lignes suivantes nous exposons les étapes utilisées pour arriver au résultat optimal.

Nous savons que la forme finale de l’équation que nous voulons estimer est la suivante :

Ainsi en réécrivant l’équation initiale notre programme d’optimisation est la suivante :

Nous posons les conditions de premier ordre (CPO) du Lagrangien.

Etant donné que est une constante, la somme allant de 1 à N est égale à

Nous pouvons extraire

Maintenant nous cherchons à exprimer

Nous remplaçons par son expression trouvée plus haut.

Or empiriquement nous savons que et sont égales à , nous allons donc multiplier le numérateur et le dénominateur par

Nous pouvous dès lors exprimer

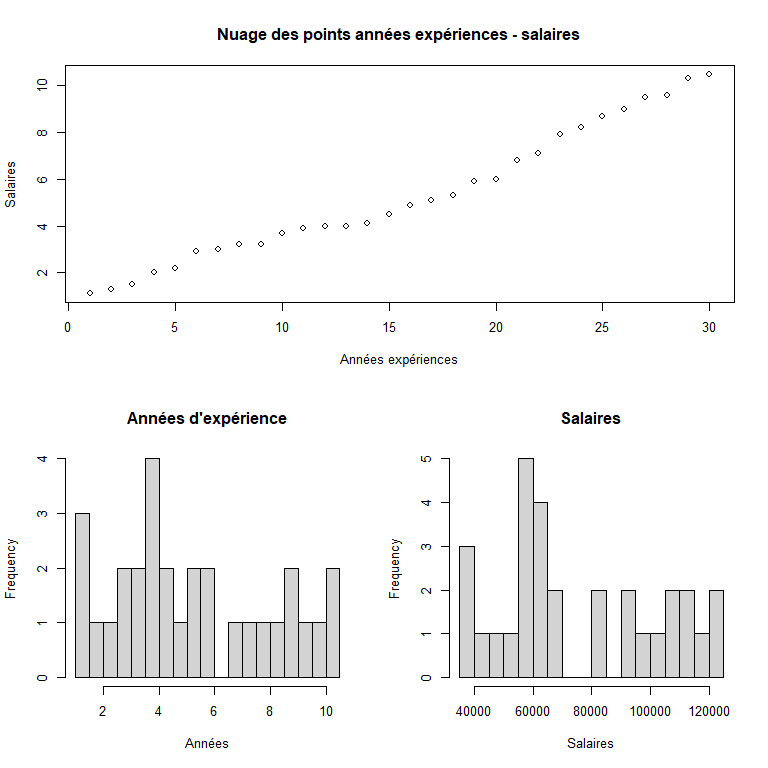
## Applications

### La relation entre le salaire et l’expérience professionnelle

Ayons un aperçu des 10 premières lignes des données sur lesquelles nous allons démontrer une application de la méthode d’estimation par les OLS simples.

|  |  |
| --- | --- |
| annees | salaire |
| 1.1 | 39343 |
| 1.3 | 46205 |
| 1.5 | 37731 |
| 2.0 | 43525 |
| 2.2 | 39891 |
| 2.9 | 56642 |
| 3.0 | 60150 |
| 3.2 | 54445 |

Nous pouvons faire des graphiques pour afficher la distribution des colonnes (par des histogrammes) et la tendance avec un graphique



Notre but est donc d’exprimer le salaire comme une fonction des années d’expérience. Nous cherchons une relation de la forme telle que les coefficients estimés et minimisent la somme du carrées des erreurs.

Nous appliquons les relations que nous avons trouvées dans lorsque nous avons résolu le programme d’optimisation du Lagrangien.

Ainsi nous calculons d’abord puis .

La formule nous indique que

|  |  |  |
| --- | --- | --- |
|  | Sommes | Moyennes |
| Années Expériences (X) | 159.4 | 5.313333 |
| Salaires (y) | 2280090.0 | 76003.000000 |

Nous allons donc calculer les statistiques intermédiaires en ajoutant des colonnes à notre tableau initial.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| 1.1 | 39343 | -4.213333 | -36660 | 154460.80 | 17.752178 |
| 1.3 | 46205 | -4.013333 | -29798 | 119589.31 | 16.106844 |
| 1.5 | 37731 | -3.813333 | -38272 | 145943.89 | 14.541511 |
| 2.0 | 43525 | -3.313333 | -32478 | 107610.44 | 10.978178 |
| 2.2 | 39891 | -3.113333 | -36112 | 112428.69 | 9.692844 |
| 2.9 | 56642 | -2.413333 | -19361 | 46724.55 | 5.824178 |

Le jeu de données contient 30 observations. Les années sont continues, il s’agit en fait du nombre total de jours qui est divisé par 365 pour exprimer les expériences en années.

## La consommation de cigarette et l’âge

• Chaque dossier devra présenter (au moins) un exemple de résolution de problème traité avec des méthode de recherche opérationnelle • Les dossiers (Powerpoint, Word, pdf, …) sont de 10 à 20 pages. Ils doivent impérativement être remis avant le 30/11/19