Examen\_mathématique

Ramya HOUNTONDJI

22/01/2021

# Introduction

L’objectif de ce rapport est de résumer puis d’évaluer Ce rapport vise à présenter et évaluer 5 travaux effectués par des étudiants de la promotion MSc Data Management 2020/2022 de PSB Paris.

L’ensemble des travaux est évalué selon les 5 critères ci-après:

* Aspect global du document ;
* Niveau de maitrise du sujet par l’auteur ;
* La qualité des codes ;
* Plue value ;
* L’intérêt intellectuel démontré ;

Je donnerais a la fin du document une autoévaluation des travaux auquels j’ai pris part.

## 1. Marche aléatoire

a- Auteur : William Robache & Marko Arsic [Github dédié à ce package](https://github.com/WilliamRbc/PSBX/tree/main/MARCHE%20ALEATOIRE)

b- Synthèse

Au travers d’un exemple de l’algorithme PageRank de Google, l’auteur nous a expliqué le concept et l’usage de la marche aléatoire. elle consiste a quantifier les pas aléatoires effectués par un agent dans un espace à dimension variée. L’ensemble des pas forment une chaîne dite ergodique qui est caractérisée par son irréductibilité, sa récurrence positive et sa non-périodicité

c- Extrait de code Explication des étapes fondamentaux de la marche aléatoire.

#Soit $d\geq 1$ et soit $(e\_1,\dots, e\_d)$ la base canonique de #$\mathbb{Z^d}$.  
#Soit $(X\_i)$ une suite de variable aléatoires indépendantes à #valeurs dans  
#$\{(\pm e\_1,\dots,\pm e\_d)\}$. On appelle marche aléatoire associée la suite  
#de variables aléatoires $(S\_n)\_{n\geq 1}$ où $S\_n$ est défini par #$$S\_n=X\_1+\dots+X\_n. $$  
#Lors de l'étude des marches aléatoires, nous souhaitons savoir si le système  
#revient à son point de départ, c'est-à-dire s'il revient à son #origine.  
#On dit que la marche alétoire est récurrente si:   
#$$ P(S\_n=0\textrm{ infiniment souvent})=1.$$  
#On dit que la marche aléatoire est transiente si:  
#$$ P(S\_n=0\textrm{ infiniment souvent})=0. $$  
#On considère une marche aléatoire sur $\mathbb Z^d$ telle que $$P(X\_i=e\_j)=P(X\_i=-e\_j)=\frac 1{2d }$$  
#pour tout $i\geq 1$ et tout $j\in\{1,\dots,d\}$.  
#Alors:   
#- La marche aléatoire est récurrente si $d=1$ ou $d=2$   
#- La marche aléatoire est transiente si $d\geq 3$

d- Evaluation

* Aspect global du document : document moyennement bien rédigé.
* Niveau de maitrise du sujet par l’auteur :sujet plus ou moins compris
* La qualité des codes : plus ou moins bien fait
* Plue value : ce sujet m’a moyennement intéressé.
* L’intérêt intellectuel démontré : l’auteur a bien montré son interet pour le travail demandé.

### Conclusion

Ce travail a permi de mettre en évidence quelques éléments mathématiques cléfs du sujet traité.

## 2.Package “dplyr`”

a- Auteur : Soukaina El Ghaldy & Jiayue Liu [Github dédié à ce package](https://github.com/soukainaElGhaldy/PSB-X/tree/main/R_packages/dplyr_package)

b- Synthèse

Les auteurs on fourni un document qui puisse permettre d’avoir une vision globale non seulement sur le package lui-même mais aussi sur son univers.Des exemples ont été donnés afi d’illustrer de façon claire les fonctions principales du package Les exemples donnés illustrent de façon claire les fonctions permises dans le package.

c- Extrait de code

Le slicing: il selectionne les lignes du tableau grace au verbe slice selon leur positions.Le premier argument va contenir le nom de la data frame et le deuxième un chiffre ou un vecteur de chiffres qui va positionner les lignes (observations) à extraire.

Le verbe slice selectionne les lignes du tableau selon leur positions.Le premier argument va contenir le nom de la data frame et le deuxième un chiffre ou un vecteur de chiffres qui va positionner les lignes (observations) à extraire.

Si on souhaite sélectionner la 10ème ligne de la table “data”:

#library(dplyr)  
#library(tibble)  
#data <- read.csv2("c:/Users/Soukaina/Documents/r\_project/test.csv")   
#data <- as\_tibble(data) #Conversion en objet tibble  
#slice(data, 10)

Les données extraites sont en format tibble.

Si on veut sélectionner uniquement les 5 premières lignes :

#slice(data, 1:5)itle = '1 mois')

d- Evaluation

* Aspect global du document : document très bien rédigé
* Niveau de maitrise du sujet par l’auteur :sujet maitrisé par l’auteur
* La qualité des codes : codes clairs et bien faits
* Plue value : ce sujet est d’un très grand interet pour
* L’intérêt intellectuel démontré : les auteurs on très bien montré son interet pour le travail demandé.

### Conclusion

Le sujet est très interessant et bien présenté mais en vérifiant on trouver quelques erreurs dans le code.

## 3.CRYPTOGRAPHIE

a- Auteur : Marko ARSIC / Wiliam ROBACHE [Github dédié à ce package](https://github.com/ARSICMrk/ARSIC_PSBx/tree/main/Maths_BD/Cryptographie)

b- Synthèse

Les auteurs de ce tutoriel nous permettent une immersion dans le monde de la cryptographie en nous faisant découvrir des algorithmes populaires tels que: le cryptage à clé symétrique et asymétrique, le système RSA, le cryptage en ligne et l’informatique quantique . Il mettent également l’accens sur leurs limites. La cryptographie est une des disciplines de la cryptologie s’attachant à protéger des messages en s’aidant souvent de secrets ou clés.

c- Extrait de code

#La confidentialité du système est basé sur la difficulté de #factoriser un nombre qui est le produit de deux grands nombres #premiers. Ce problème fait effectivement parti des problèmes dits #difficiles en mathématiques.   
  
#La fonction de cryptage est l’exponentation à la puissance $e$ modulo $n$, où $e$ est appelé clé d’encryptage et $n$ est le modulo du cryptage (il s’agit d’un cryptage par bloc).   
  
#La fonction de décryptage est l’exponentation à la puissance  
#$d$ modulo $n$, où $d$ est appelé clé de décryptage.   
  
#Pour que le système fonctionne, il faut que $n$ soit choisi de  
#façon à être le produit de deux nombres premiers $p$ et $q$.   
  
#Ensuite $e$ et $d$ doivent être choisis de façon à ce que $$e.d\equiv  
#1[(p − 1)(q − 1)]$$

d- Evaluation

* Aspect global du document : document bien rédigé
* Niveau de maitrise du sujet par l’auteur :sujet très bien maitrisé par les auteurs.
* La qualité des codes : codes clairs et bien faits
* Plue value : ce sujet est d’un très grand interet meme si j’aurais aimé plus voir les aspects mathématiques.
* L’intérêt intellectuel démontré : l’auteur a bien montré son interet pour le travail demandé.

### Conclusion

Le travail a été bien fait et expliqué mais il aurait été intéressant d’avoir les formules de cryptage et de décryptage proprement dites sous forme mathématique.

## 4. “DECISION TREES”

a- Auteur : BRETONNIER Corentin / Benjamin GUIGON / Antoine SERREAU [Github dédié à ce package](https://github.com/aserreau/PSB1/tree/main/Travaux%20Mathématiques)

b- Synthèse

Les arbres de decisions modelisent une hierarchie de test pour prendre une decision ou predire un resultat en fonction des experiences precedentes. Il en existe deux types : - Les arbres de r´egressions (Regression Tree) permettent de predire une reponse quantitative (par exemple le prix d’une maison). - Les arbres de classifications (Classification Tree) permettent de predire une reponse qualitative (Identification d’une maladie et de ses symptomes. c- Extrait de code

d- Evaluation

* Aspect global du document : globalement le document est bien rédigé, on observe une partie du code pas très lisible.
* Niveau de maitrise du sujet par l’auteur :sujet maitrisé par l’auteur et contenu très fourni
* La qualité des codes : formules mathématiques pas évidentes a appréhender.
* Plue value : ce sujet est d’un très grand interet pour différents secteurs comme la médecine et la finance.
* L’intérêt intellectuel démontré : l’auteur a bien montré son interet pour le travail demandé.

### Conclusion

Le travail des auteurs a une réelle plus value. De plus le fait qu’il nous orientent vers un autre camarade permet de voir qu’ils ont bien de la suite dans leurs iddées.

## 5. “svm”

a- Auteur : Ramya HOUNTONDJI [Github dédié à ce package](https://github.com/RamyaHTDJ/Psb_Ramya/blob/main/svm.pdf)

b- Synthèse

Une machine a vecteurs de support, traduction litterale pour Support Vector Machine, est un algorithme d’apprentissage automatique supervisé qui peut etre utilise a des fins de classification et de regression. Les SVM sont plus generalement utilises dans les situations de classification. Ils ont été d´eveloppés dans les années 1990. Leur principe est simple : il ont pour but de séparer les données en classes a l’aide d’une frontiere aussi simple que possible, de telle façon que la distance entre les différents groupes de données et la frontière qui les sépare soit maximale. Cette distance est aussi appelée marge et les SVMs sont ainsi qualifiés de séparateurs a vaste marge , les vecteurs de support étant les données les plus proches de la frontière.

d- auto Evaluation

* Aspect global du document : le document a été rédigé simplement et avec latex pour les formules..
* Niveau de maitrise du sujet par l’auteur :j’ai globalement bien compris le sujet.
* La qualité des codes : peu de code.
* Plus value : ce sujet est d’un très grand interet pour faire de la classification de données pas trop volumineuses.
* L’intérêt intellectuel démontré : Nous avons bien montré l’interet du sujet.