
title: "Partiel" author: "Nicolas Allix"

output: pdf_document

Math

Modèl de régression

Résumé

[Lien](#)

Il existe deux types de régressions, simple et multiple. Les deux sont modélisés sur des bases statistiques. La régression linéaire simple nous permet d'établir un lien entre deux variables quantitatives. On modélise deux variable autour d'une ligne droite. Si les points se rassemble, alors il y'a une relation linéaire entre les deux variables. Il est aussi possible de modéliser la relations avec des relations quadratique, cubique ... Quand l'hypothèse est nulle, il n'y a pas de relation entre les variables. Pour le modéliser en R, il faut d'abord selectionner les deux variables d'une base de donnée.

```
{r} y=g$Age x=g$Elevel plot(y,x)
```

 Après avoir défini les données, il faut exécuter le plot pour visualiser la relation.

La régression multiple suit la même logique que la régression simple, mais elle a plusieurs variables explicatives

Note

L'explication est clair avec des exemples concrets pour mieux comprendre l'utilisation d'un modèl de régression. Le code est assez expliqué avec plusieurs modèle possible.

Clarté : 5/5

Expression mathématique : 4/5

Vulgarisation : 5/5

Propreté du document : 5/5

Originalité : 4/5

Cryptographie et théorie des nombres

Résumé

[Lien](#)

La cryptographie a été développé au milieu du 20ème siècle dans un premier temps pour le domaine militaire puis a trouvé des utilisations dans le domaine civile. Il y a deux type de cryptographie. Le cryptage à clé symétrique avec le codage César et le codage Vigènère. Il y a aussi le cryptage asymétrique avec la fonction à sens unique et la fonction à sens unique avec Trappe.

Note

Le thème étudié est très interessant, la vulgarisation de de la cryptographie est très interessante de plus que c'est un domaine qui a de plus en plus d'importance avec la quantité de donnée qui circule aujourd'hui. L'explication des fonctions utilisé pour le chiffage était très clair.

Clarté : 7/5

Expression mathématique : 4/5

Vulgarisation : 6/5

Propreté du document : 5/5

Originalité : 5/5

Forecasting

Résumé

[Lien](#)

Ce travail nous explique ce qu'est prédiction à l'aide de trois articles. Qu'est-ce qu'une bonne prédiction ? / Peut-on prédire le décrochage des personnes qui suivent les MOOCs ? / Peut-on prédire les étudiants à risque avec leur comportement d'apprentissage online et offline ? Dans leur cas, ils s'intéresse notamment au sujet du décrochage scolaire en fonction du comportement numérique de ce dernier. Ils font référence à la formule du GVF. Les GVF peuvent être interprétées comme un rendement attendu à une action C. Ils font notamment référence au mooc et aux algorithmes de branche, de liaison modifié et MLP pour relier l'activité numérique au décrochage. Enfin ils ont parlé de l'algorithme EPARS pour prédire les risques d'abandon de cours.

Note

Très clair et intéressant, ils ont rendu les articles plus compréhensible.

Clarté : 5/5

Expression mathématique : 5/5

Vulgarisation : 5/5

Propreté du document : 5/5

Originalité : 5/5

Convolutional Neural Network (CNN)

Résumé

[Lien](#)

Le réseau de neurones convolutifs spécifie une sous-catégorie de réseau de neurones et est l'un des modèles de classification d'images les plus réussis au monde à ce jour. À première vue, leur mode de fonctionnement est simple: l'utilisateur fournit une image en entrée sous la forme d'une matrice de pixels. La disponibilité de grandes quantités de données et les améliorations de la technologie matérielle Des recherches accélérées sur CNN et des architectures CNN profondes intéressantes sont récemment apparues. Il est ensuite expliqué toutes les différentes couches d'une architecture CNN

Note

Très intéressant et bien vulgarisé.

Clarté : 5/5

Expression mathématique : 5/5

Vulgarisation : 5/5

Propreté du document : 5/5

Originalité : 5/5

MARCHE ALEATOIRE

Résumé

[Lien](#)

Les étapes de la marche aléatoire sont totalement indépendantes les unes des autres. Intuitivement, cela signifie que l'avenir du système dépend à tout moment de son état actuel, pas de son passé, ni même de son approche la plus proche. En d'autres termes, le système «perd de la mémoire» avec le temps. (Une autre raison pour "l'homme ivre se promenant") Google utilise des marches aléatoires pour parcourir, identifier et classer les pages Internet. La marche aléatoire est décomposée en une unité de base appelée *pas*, et sa longueur elle-même peut être constante, aléatoire ou fixée par le réseau ou les graphiques que nous voulons parcourir. Par conséquent, à chaque étape, nous avons une gamme de possibilités pour choisir au hasard la direction et la taille de l'étape. Cette gamme de possibilités peut être discrète (choisir parmi un nombre limité de valeurs) ou continue.

Note

Travail très clair avec des formules bien expliqué et une théorie bien vulgarisée.

Clarté : 5/5

Expression mathématique : 5/5

Vulgarisation : 5/5

Propreté du document : 5/5

Originalité : 5/5

R(rrr)

R.miner

Résumé

[Lien](#)

Rminer est un package R qui permet de faire du data mining à l'aide d'algorithmes. Il permet notamment de faire de la classification et de la régression. On peut répartir les fonctions de ce package en 3 parties. La préparation de donnée avec les fonctions CasesSeries et delevels. Le modelage de la donnée avec les fonctions fit, predict et modeling. Et enfin l'évaluation du modèle avec les fonctions mmetric, mgraph, mining.

Note

Le contenu du package a été bien explicité. Avoir catégorisé les fonctions permet de mieux comprendre quand les utiliser et pourquoi.

Clarté : 5/5

Expression R : 5/5

Vulgarisation : 5/5

Propreté du document : 5/5

Originalité : 5/5

MNIST et Fashion MNIST (Illustration dans R)

Résumé

[Lien](#)

MNIST et Fashion MNIST sont les bases de l'apprentissage supervisé. MNIST est un jeu de données très utilisé en apprentissage automatique. Fashion MNIST contient plus de 70000 images d'habilles. Elles permettent d'entraîner les algorithmes de machine Learning.

Note

Le contenu du package a été bien explicité. C'est un parfait début pour se lancer dans l'apprentissage automatique.

Clarté : 5/5

Expression R : 5/5

Vulgarisation : 5/5

Propreté du document : 5/5

Originalité : 5/5

Data visualisation interactive avec Plotly

Résumé

[Lien](#)

Le package Plotly nous permet de créer des graphiques interactifs de différentes qualités. En utilisant ce package, vous pouvez créer des graphiques linéaires, des nuages de points, des graphiques à barres, des boîtes à moustaches, des histogrammes, des cartes et d'autres graphiques. Il existe deux manières de créer un objet plotly. En transformant un objet ggplot2 (via `ggplotly()`), et en initialisant directement un objet plotly avec `plot_ly()` / `plot_geo()` / `plot_mapbox()`. On voit aussi toutes les fonctions en détaille pour créer des visualisations percutantes.

Note

Présentation très complète du package avec toutes les possibilités existantes.

Clarté : 5/5

Expression R : 5/5

Vulgarisation : 5/5

Propreté du document : 5/5

Originalité : 5/5

Rstatix

Résumé

[Lien](#)

Le package "Rstatix" fournit un cadre simple et intuitif pour effectuer des tests statistiques de base (y compris le test de comparaison des moyennes, le test de Wilcoxon, l'ANOVA, Kruskal-Wallis et l'analyse de corrélation). Ils ont utilisé un jeu de données sur l'estime de soi mesurée sur trois points temporels fourni par le package "datarium". Grâce à ce jeu de données, ils ont effectué les tests en les expliquant.

Note

Peut être reproduit pour pouvoir travailler dessus en autonomie. Clarté : 5/5

Expression R : 5/5

Vulgarisation : 5/5

Propreté du document : 5/5

Originalité : 5/5

GGPLOT 2

Résumé

[Lien](#)

GPLOT2 est une bibliothèque R qui vous permet de visualiser les données développées par Hadley Wickham. GGPLOT 2 est une extension de tidyverse, qui permet la génération de graphiques en utilisant une syntaxe cohérente et puissante. Une des caractéristiques de GGPLOT 2 est qu'il suppose que les données liées au graphique sont stockées dans une table de données (bloc de données ou autre). Ils ont ensuite réaliser l'analyse d'un jeu de données de 20 états occidentaux afin de bien comprendre comment le package fonctionne et facilement être autonome dessus.

Note

Permet de facilement prendre en mains le package avec un exemple clair, c'est génial !

Clarté : 5/5

Expression R : 5/5

Vulgarisation : 5/5

Propreté du document : 5/5

Originalité : 5/5

Mes travaux

Rattle

[Lien](#)

Le travail sur Rattle était très intéressant. j'ai pu apprendre ce qu'est un GUI mais plus important encore, j'ai découvert le Package magique de R pour faire du Machine learning rapidement et simplement. la prise en main de l'outil n'est pas trop compliqué, il suffit juste d'avoir les bases de Machine learning pour comprendre comment utiliser Rattle. Dans ce travail, j'ai passer un peut trop de temp à expliquer une grande partie des fonctions Rattle, et pas assez sur comment bien utiliser l'outils. Nous avons néanmoins ajouté une vidéo tutoriel qui montre plus en détail l'interface de Rattle. Ce qu'il manque à ce travail, c'est de voire tout le code qui a été généré pendant le tutoriel.

Les arbres de décision

[Lien](#)

Ce travail est bien organisé. Nous sommes partis des informations les plus générales au plus précise en allant à chaque fois un peu plus en profondeur afin de comprendre brique par brique ce qu'est un arbre de décision. Nous avons bien expliqué globalement à quoi servent les formules et comment lire les résultats. Il aurait peut être fallu faire une étude de cas avec un exemple plus concret afin d'avoir un travail plus original.