

AstéRix_teddy_leandre_caret

Teddy Leandre

23/12/2020

I. Présentation

Dans ce dossier, nous chercherons à évaluer 12 travaux comportant du code R et des concepts mathématiques. L'évaluation se fera sur la base de 5 critères de notation qui, additionnés, donnera une note globale sur 20 points.

1. Qualité du visuel du document sur 4 points
2. Pluralité des fonctionnalités sur 4 points
3. Fonctionnement du code sur 4 points
4. Qualité des explications sur 4 points
5. Lisibilité du code sur 4 points

II. Description

Ici nous verrons l'outil `caret`, c'est un paquet qui permet d'effectuer de l'analyse prédictive. Le travail de Teddy Leandre (c'est bien moi :)) se porte sur certaines des fonctionnalités principales du paquet.

[Le Github évalué](#)

a. Commentaire

- `Caret` est un paquet de R qui permet le traitement de donnée dans le cadre du machine learning. Il permet la classification ainsi que la régression de modèle. La construction de modèles pour un dataset, certaines tâches sont tout aussi importantes que l'élaboration de l'algorithme du machine learning lui-même. On peut notamment noter le nettoyage des données, le traitement d'observations incomplètes ainsi que la validation du modèle sur un ensemble de tests. *

II) Cas pratiques

On utilisera le dataset `iris` que l'on traitera et on téléchargera le paquet `caret` ainsi que le paquet `ggplot2` et `lattice` qui permettront a certaines fonctionnalités de marcher. On entamera ensuite étape par étape, les démarches pour créer un modèle.

On va configurer ce modèle prédictif en téléchargeant les paquets et le jeu de données

```
# Téléchargement des différents paquets
install.packages("caret")
install.packages("ggplot2")
install.packages("lattice")
# Importation de la table de donnée `iris`
iris
# Chargement du paquet caret
library(caret)
# Structure de la table de données
str(iris)
# Observation des 6 premières lignes du tableau
head(iris)
```

Préparation des données et pré-traitement

```
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
# Création des ensembles de données d'entraînement et de test
set.seed(100)
# Etablir un nouvel index pour les chaque ligne `species`
iris_index <- createDataPartition(iris$Species, p = .8,
                                  list = FALSE,
                                  times = 1)

# Creation d'un dataset d'entraînement
Train_iris <- iris[iris_index,]
# Création d'un dataset sur lequel se feront les tests
Test_iris <- iris[-iris_index,]
```

“La science, aujourd’hui, cherchera une source d’inspiration au-dessus d’elle ou périra.” **Simone Weil**

C’est pourquoi dans le cadre de mon mini tutoriel je tiens à citer mes sources : [Source 1](#) || [Source 2](#)

Notation

Critère 1 : 3/4 Peut être amélioré.

Critère 2 : 3/4 Utilisation de fonctions essentiels mais reste simple.

Critère 3 : 4/4 tout fonctionne.

Critère 4 : 3/4 Explications brief sur certaines lignes, . Des commentaires supplémentaires peuvent être ajoutés.

Critère 5 : 4/4 très simple et lisible.

III. Conclusion

Note globale 17/20. Les fonctionnalités du paquet présentés sont lisibles et présentables. Les sources sont bien citées