

Le travail de Teddy Leandre est disponible ici

BENOUCIEF Amine

10/12/2020

SYNTHESE DU TRAVAIL EN QUESTION

Le package “caret” (Classification And REgression Training) est une librairie pour R. Il couvre une large fraction de la pratique de l’analyse prédictive (classement et régression). Un peu à la manière de “scikit-learn” pour Python, il intègre dans un ensemble cohérent les étapes clés de la modélisation : préparation des données, sélection, apprentissage, évaluation. Teddy a bien montré qu’il sait utiliser le package caret. Chaque ligne est bien expliquée et donc pas nécessaire d’ajouter de commentaires.

I) Introduction

- Caret est un paquet de R qui permet le traitement de données 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 certaines fonctionnalités de marche. 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 données `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)  
  
## Warning: package 'caret' was built under R version 4.0.3  
## Loading required package: lattice  
## Warning: package 'lattice' was built under R version 4.0.3  
## 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 :

caption [Source 1](#) || [Source 2](#)

Caption [Mon Github](#)

Revenir à [L’introduction](#)

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

EVALUATION DU TRAVAIL EN QUESTION

Critère 1 : Visuel sur pdf 3/4 Très Agréable à lire.

Critère 2 : Originalité du code 2/4 Démarche simple.

Critère 3 : Fonctionnalité du code 4/4 le code fonctionne.

Critère 4 : Lisibilité du code 4/4 Claire et lisible.

Critère 5 : Explications données 4/4 Toutes les explications sont claires sur chaque ligne de code.

CONCLUSION

Globalement un bon travail qui explique très bien le projet. Intégration d'une analyse prédictive pour compléter le travail.