

Analyse de caret par Jean Souris

Wenjun ZHAO

12/20/2020

I. Introduction

Vous pouvez retrouver mon travail sur mon Github, <https://github.com/claudia0524/PSBX>.

Pour ce dossier, nous avons étudié le travail de Jean Souris, étudiant en M2DM au sein de PSB.

Nous avons trouvé son dossier sur son Github : <https://github.com/jeansouris/PSBX.git>

Maintenant, nous allons établir mes 5 critères d'évaluations, qui seront les mêmes pour tous les dossiers que j'ai vus étudier pour ce devoir:

Les cinq critères d'évaluation

1. *Rmd se comporte bien à l'exécution*
2. *Les aspects intéressants, didactiques, complets*
3. *La qualité Rmarkdown, la qualité de l'écriture*
4. *Didactique, conformité aux exigences vues plus haut et comporte du calcul symbolique et du calcul numérique*
5. *La qualité du LaTeX et des illustrations, la qualité de l'écriture, le choix des ressources internet, la compréhension personnelle des concepts*

II. Synthèse du travail en question

Caret est utilisé pour le Machine Learning.

Caret est un des plus gros packages de R, qui permet de faire de la prédiction de données. Il y a 2 types de modèles que nous pouvons prédire avec Caret : la Classification, qui relève le côté qualitatif, et la régression, qui elle se concentre sur le côté quantitatif.

Dans ce dossier, nous allons utiliser une fonction qui nous montrera la séparation des données d'une data frame.

Mais, avant de passer à l'application, nous allons avoir besoin des packages "lattice", "ggplot2", et "Caret", afin d'exploiter les données que nous souhaitons (Souris, n.d.).

Nous allons maintenant étudier le code de Jean dans le chapitre suivant.

III. Extrait commenté des parties de code

```
#install.packages("lattice")  
#Ensuite, nous allons les appeler avec la fonction library :  
library(lattice)  
library(ggplot2)  
library(caret)
```

```

#Création d'un tableau séparant les données en gardant l'écart-type global : createDataPartition
#Tout d'abord, nous allons générer un nombre aléatoire grâce à la fonction "set.seed" :
set.seed(2020)

#Maintenant que notre chiffre aléatoire est généré, nous pouvons créer nos différents tableaux grâce à
Essaie_1 <- createDataPartition(iris$Species, p = .6,
                                list = FALSE,
                                times = 3)

head(Essaie_1)

##      Resample1 Resample2 Resample3
## [1,]         1         2         1
## [2,]         2         4         2
## [3,]         3         7         4
## [4,]         4         8         7
## [5,]         6         9         8
## [6,]         7        13         9

#Notre tableau a donc généré des nombres aléatoire entre les différentes espèces d'iris. Dans cet exemp
#Le terme "liste = FALSE" permet d'éviter que les données ne se retrouve sous forme de liste au lieu de
Essaie_2 <- createDataPartition(iris$Species, p = .6,
                                list = TRUE,
                                times = 2)

head(Essaie_2)

## $Resample1
## [1]  3  4  6  7 10 11 12 14 15 17 18 19 20 25 26 28 30 31 32
## [20] 33 36 37 38 42 43 44 45 47 48 50 52 55 57 58 60 64 65 67
## [39] 69 71 72 73 74 76 77 78 79 80 81 82 84 86 89 90 92 93 96
## [58] 97 98 99 101 105 106 108 110 112 113 115 116 118 119 120 121 122 123 124
## [77] 125 128 129 130 132 134 135 137 138 141 144 147 148 150
##
## $Resample2
## [1]  1  3  4  6  7  9 10 11 12 14 15 17 19 20 21 24 26 27 28
## [20] 29 30 33 36 37 38 42 43 44 45 47 52 53 56 57 58 60 62 65
## [39] 69 70 72 76 77 80 81 82 83 85 87 89 90 91 92 93 94 95 96
## [58] 98 99 100 102 103 104 106 107 108 109 110 111 113 115 116 118 119 122 123
## [77] 125 126 127 128 129 134 136 138 139 141 143 145 146 150

```

IV. Evaluation du travail en question

1. Rmd se comporte bien à l'exécution (4/4)
2. Les aspects intéressant, didactique, complet (3/4)
3. La qualité Rmarkdown, la qualité de l'écriture (4/4)
4. Didactique, conformité aux exigences vues plus haut et comporte du calcul symbolique et du calcul numérique (3/4)
5. La qualité du LaTeX et des illustrations, la qualité de l'écriture, le choix des ressource internet, la compréhension personnelle des concepts (3/4)

V. Conclusion

En général, ce tracail de Jean exécute bien dans l'environnement de R. Et il nous explique clairement comment fonctionner le package Caret dans un rmd, On comprend rapidement le but principal de ce package. D'ailleurs, la qualité de Rmarkdown et l'écriture sont ainsi bien, conformité aux exigences vues plus haut. L'aspect est intéressant, complet et propre. La qualité Rmarkdown et la qualité de l'écriture sont plus hauts, La qualité du LaTeX et des illustrations ne sont pas mal.

VI. Bibliographie

Souris, Jean. n.d. "Travail de Caret." <https://github.com/jeansouris/PSBX.git>.