

Résumé_R_Caret_Jean

Xueting YIN

2020/12/23

Entête :

Nous allons analyser le package caret dans le langage R expliqué par Jean SOURIS, le lien vers son github est <https://github.com/jeansouris/PSBX>.

Synthèse :

Dans ce document, Jean a choisi le data frame iris, il a chargé trois packages : lattice, ggplot2, caret pour nous montrer comment extraire un certain pourcentage de données dans le groupe d'entraînement et le groupe de test avec la fonction createDataPartition(), il a comparé les résultats de deux extractions et de trois extractions afin que les lecteurs puissent voir la différence en un coup d'œil.

Extrait commenté :

Nous allons commenter la partie « Création d'un tableau séparant les données en gardant l'écart-type global : createDataPartition ». Jean nous guide pas à pas comment extraire une certaine proportion de données à travers un récit textuel. D'abord il faut générer un nombre aléatoire grâce à la fonction set.seed():

```
library(lattice)
library(ggplot2)
library(caret)
set.seed(2020)
```

Ensuite, il nous montre les résultats de l'extraction de soixante pour cent des données à trois reprises et explique en détail la signification de chaque ligne de code, par exemple ce que signifie .6 et pourquoi la liste est fausse, et enfin, montrent ce qui se passe si vous n'extrayez que deux fois soixante pour cent des données :

```
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

```
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
```

L’explication de Jean de cette partie d’extraction de données est relativement complète.

L’évaluation :

1. Le document est-il visuellement agréable ?

L’ensemble du document est visuellement confortable, avec une mise en page claire, des sous-titres clairs, concis et plus propices à la compréhension et à la lecture.

2. Est-il compréhensible ? (qualité des explications)

Comme le package caret dans R a de nombreuses fonctions, Jean prend l’une des plus importantes comme point d’entrée et nous montre très bien la fonction createDataPartition() avec un langage explicatif et une présentation des résultats, de sorte que l’ensemble du document est compréhensible.

3. Combien de fonctions ont été utilisé ?

Jean a utilisé principalement deux fonction : set.seed() pour générer un nombre aléatoire, createDataPartition() pour extraire un pourcentage défini des données.

4. Est-il intéressant d'utiliser telles fonctions ?

Oui, caret est un package pour construire un modèle prédictif, mais avant de construire ce modèle, nous avons besoin deux groupes de données, donc ici Jean nous a montré la fonction importante dans le package caret.

5. Y-a-t-il une bibliographie?

Jean n’a pas mis un titre bibliographie, mais il a cité une source pour la remercier et pour que les lecteurs puissent savoir plus d’informations sur ce package.

Conclusion :

J’apprécie la partie expliquée par Jean dans un langage concis et clair, qui permet aux lecteurs de mieux comprendre le contexte de fonctionnement de ce package. De plus, j’apprécie également la partie de comparaison, qui nous offre de nouvelles options. Bien sûr, le défaut de ce document est qu’il est trop succinct et nous pouvons connaître que deux fonctions.