

Caret

Jean Souris

15/12/2020

```
#install.packages("caret")
```

```
library(caret)
```

Caret est utilisé pour le Machine Learning.

Caret est un des plus gros package de R, qui permet de faire de la prediction de données. Il y a 2 types de modèle que nous pouvons prédire avec Caret : la Classification, qui relève le côté qualitatif, et la regression, 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 :

```
#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 la fonction "createDataPartition" :

```
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 exemple, nous avons séparé les données de la manière 60% - 40%. Le terme "liste = FALSE" permet d'éviter que les données ne se retrouve sous forme de liste au lieu de tableau. (comme vous pourrez le voir ci-dessous) :

```
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
```

Merci de votre lecture !

Retrouvez tous mes dossiers juste ici : [Mon Github](#)

Et merci à cette source de m'avoir aidé ! [Source](#)