

Résumé_R_Cross validation_Marko/Rindra

Xueting YIN

2020/12/23

Entête :

Nous allons analyser un document de cross validation expliqué par Marko ARSIC / Rindra LUTZ, le lien vers leur github est https://github.com/ARSICMrk/ARSIC_PSBx/tree/main/R_Travail_Sup.

Synthèse :

Ce document est basé sur le cross validation, mais Marko et Rindra nous ont expliqué un peu la situation maintenant dans le domaine data, ils nous ont présenté également la méthode descriptive et la méthode prédictive, c'est à partir de cette méthode prédictive que nous avons le cross validation qui est une régression logique, et aussi la partie principale dans ce document, ils nous ont montré quelques exemples pour nous faire comprendre mieux ce cross validation.

Extrait commenté :

Nous allons commenter l'exemple utilisé dans ce document. Marko et Rindra ont d'abord installer et charger le package tidyverse et caret qui est indispensable pour construire un modèle prédictif. Ensuite ils utilisent le data frame swiss pour effectuer les analyses.

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## √ ggplot2 3.3.2   √ purrr  0.3.4
## √ tibble 3.0.4    √ dplyr  1.0.2
## √ tidyr  1.1.2    √ stringr 1.4.0
## √ readr  1.4.0    √ forcats 0.5.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(caret)
```

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

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
## lift
```

```
data("swiss")
sample_n(swiss, 3)
```

```
##           Fertility Agriculture Examination Education Catholic
## St Maurice      65.0          75.9           9           9    99.06
## Delemont        83.1          45.1           6           9    84.84
## Lausanne        55.7          19.4          26          28    12.11
##           Infant.Mortality
## St Maurice              17.8
## Delemont                22.2
## Lausanne                20.2
```

Et puis, ils utilisent la méthode cross validation pour définir l'échantillon d'entraînement.

```
set.seed(123)
train.control <- trainControl(method = "cv", number = 10)
```

Enfin, ils ont entraîné le modèle :

```
model <- train(Fertility ~., data = swiss, method = "lm",
               trControl = train.control)
print(model)
```

```
## Linear Regression
##
## 47 samples
## 5 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 42, 42, 42, 42, 42, 44, ...
## Resampling results:
##
##      RMSE      Rsquared    MAE
## 7.424916 0.6922072 6.31218
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

Pendant tout l'exemple, les auteurs ont très bien expliqué le processus et chaque ligne de code, ils ont même comparé juste après les résultats en utilisant le cross validation avec 3 répétitions .

Evaluation :

1. Le document est-il visuellement agréable ?

Il est très agréable visuellement, parce que les sous-titres de chaque section sont très clairs, qu'il y a des images pour expliquer les descriptions et que les couleurs distinguent les points clés.

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

Il est parfaitement compréhensible car les auteurs utilisent de nombreux énoncés explicatifs qui détaillent pourquoi et comment utiliser la validation croisée, en menant pas à pas.

3. Combien de fonctions ont été utilisé ?

Ils ont utilisé beaucoup de fonctions principalement dans le package caret pour cette partie de cross validation.

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

Oui, la méthode cross validation est très souvent utilisé pour construire un modèle prédictif, donc c'est très intéressant pourquoi et comment utiliser cette méthode.

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

Oui, il y a une bibliographie pour trouver plus d'information.

Conclusion :

J'apprécie vraiment l'exhaustivité de ce document, les auteurs commencent par un historique de la méthode et vont étape par étape afin que nous, les lecteurs, puissions pleinement comprendre pourquoi et comment utiliser cette méthode de validation croisée. Grâce à ce document, nous pouvons totalement comprendre cette méthode sans avoir aucun problème.