

Introduction à la crossvalidation

Priestley Muhindo

2021-05-31

Contents

1	Prérequis	5
2	Importation de Données	7
2.1	Chargement des libraries	7
2.2	importation des données	7
2.3	Visualisation des données	7
3	Régression Linéaire	9
4	Partitionnement (rééchantillonnage)	11
5	Cross-validation	13
6	Meilleurs hyperparamètres	15

Chapter 1

Prérequis

Chapter 2

Importation de Données

2.1 Chargement des libraries

```
library(tidymodels)

library(readr)
library(broom.mixed) # convertir les résultats en tibble
library(dotwhisker) # Visualisation de données
library(forcats)
```

2.2 importation des données

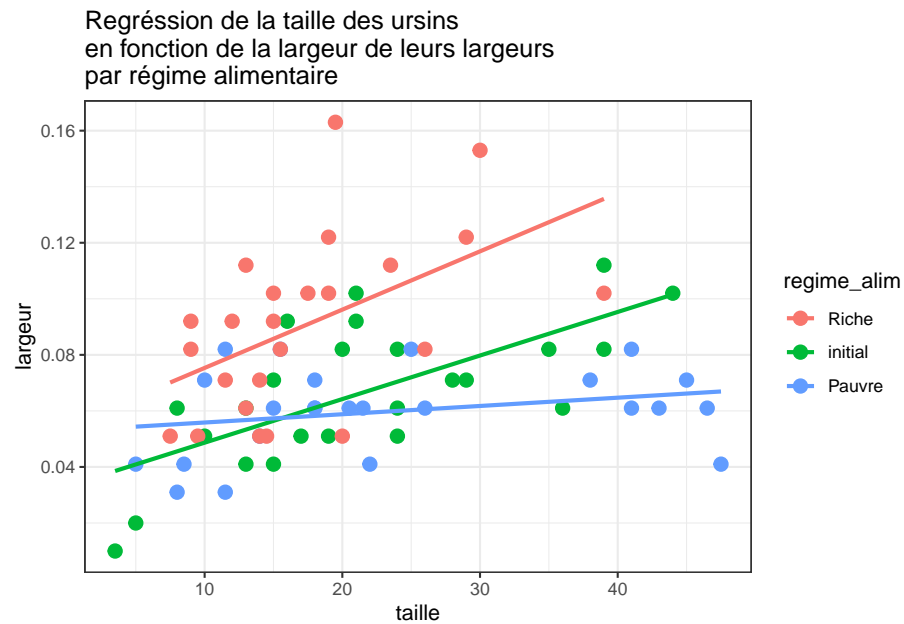
Les données des oursins issues d'une expérience au laboratoire

```
urchins <- read_csv("https://tidymodels.org/start/models/urchins.csv") %>% setNames(c("regime_alim", "taille", "largeur", "profondeur", "sexe", "annee"))
mutate(regime_alim=factor(regime_alim))
urchins$regime_alim <- fct_recode(urchins$regime_alim ,
                                "initial" = "Initial",
                                "Pauvre" = "Low",
                                "Riche" = "High")
```

2.3 Visualisation des données

```
ggplot(urchins,
       aes(x=taille, y=largeur, group=regime_alim, color=regime_alim)) +
  geom_point(size=3) +
  geom_smooth(method = "lm", se=F) +
```

```
theme_bw()+
labs(title = "Régression de la taille des ursins \nen fonction de la largeur de leurs",
```



On peut vite remarquer que que la différence se dégage entre les différents régimes alimentaires

Chapter 3

Régression Linéaire

Nous allons utiliser un modèle linéaire avec variables qualitatives en mettant en exergue l'interaction car nous avons à la fois un prédictor continu et un prédictor catégoriel. Étant donné que les pentes semblent être différentes pour au moins deux des régimes d'alimentation, construisons un modèle qui permet des interactions bidirectionnelles. Spécifier une formule R avec nos variables de cette manière : Formule du modele et package

```
lm_mod <-  
  linear_reg() %>% # Précision du modèle  
  set_engine("lm") # Précision du package
```


Chapter 4

Partitionnement (rééchantillonnage)

Chapter 5

Cross-validation

Chapter 6

Meilleurs hyperparamètres