

---

# **Fonctions Regression**

***Version 2023***

**Benali Nafissa et Fuentes Vicente Laura**

**nov. 09, 2023**



---

## Contents:

---

<b>1</b>	<b>combinaison module</b>	<b>1</b>
<b>2</b>	<b>dataset_division module</b>	<b>3</b>
<b>3</b>	<b>general module</b>	<b>5</b>
<b>4</b>	<b>Indices and tables</b>	<b>7</b>
	<b>Index des modules Python</b>	<b>9</b>
	<b>Index</b>	<b>11</b>



# CHAPITRE 1

---

combinaison module

---



---

### dataset\_division module

---

`dataset_division.formal_div(data, idx0, idx1)`

Fonction qui sépare le jeu de données data en deux jeux de données data0 (indices où wine\_type=0) et data1 (resp. wine\_type=1)

**Paramètres**

- **data** (*pd.DataFrame*) – Jeu de données à séparer en data0 et data1
- **idx0** (*np.array*) – Indices où wine\_type=0
- **idx1** (*\_type\_*) – Indices où wine\_type=1

**Renvoie**

Jeu de données contenant les vins de type 0 data1 (*pd.DataFrame*) : Jeu de données contenant les vins de type 1

**Type renvoyé**

data0 (*pd.DataFrame*)

`dataset_division.treatment(data)`

Fonction qui sépare un dataset donné en co-variables (X) et variable explicative (y). NB : cette fonction sera implémentée que sur les jeu de données train (data0, data1)

**Paramètres**

**data** (*pd.DataFrame*) – Jeu de données à séparer X, y

**Renvoie**

Co-variables du jeu de données y (*pd.DataFrame*) : Variable explicative

**Type renvoyé**

X (*pd.DataFrame*)

`dataset_division.winetype(data)`

Fonction qui repère les indices d'un dataset où la variable « wine\_type » est 0 et 1.

**Paramètres**

**data** (*pd.DataFrame*) – Jeu de données à l'origine de la division en fonction de la variable « wine\_type »

**Renvoie**

vecteur contenant les indices des lignes telles que « wine\_type »=0 *index\_1* (*np.array*) : vecteur contenant les indices des lignes telles que « wine\_type »=1

**Type renvoyé**

index\_0 (np.array)



`general.CV_rep(Xtr, ytr, nfolds)`

La fonction crée des nouveaux jeux de données en sous-divisant les jeux de données en

**Paramètres**

- **Xtr** (*pd.DataFrame*) – Jeu de données à diviser, contenant les co-variables
- **ytr** (*pd.DataFrame*) – Vecteur à diviser, contenant la variable à prédire
- **nfolds** (*int*) – Nombre représentant en combien de sous dataframes on souhaite diviser Xtr et ytr

**Renvoie**

liste contenant (n-folds) jeux de données y\_new (list) : liste contenant (les n-folds) nouvelles version de ytr

**Type renvoyé**

X\_new (list)

`general.build_pred(X_test0, X_test1, pred0, pred1)`

Fonction qui combine les prédictions de data0 et data1

**Paramètres**

- **X\_test0** (*pd.DataFrame*) – Vecteur contenant les co-variables à wine\_type=0 du jeu de données test
- **X\_test1** (*pd.DataFrame*) – Vecteur contenant les co-variables à wine\_type=0 du jeu de données test
- **pred0** (*np.array* ou *pd.DataFrame*) – prédictions associés pour les indices de wine\_type=0
- **pred1** (*np.array* ou *pd.DataFrame*) – prédictions associés pour les indices de wine\_type=1

**Renvoie**

Vecteur contenant l'ensemble des prédictions du jeu de données test

**Type renvoyé**

pred (*np.array*)

`general.param_selection(param, mod, Xtr, ytr, Xte)`

\_summary\_

### Paramètres

- **param** (*liste*) – { « param1 » : [“option1”, “option2”], » param2 » : [option1, option2, option3]}
- **mod** (*sklearn.function*) – model(random\_state=10)
- **Xtr** (*pd.DataFrame*) – Co-variables du jeu de données train
- **ytr** (*pd.DataFrame*) – Variable à prédire du jeu de données train
- **Xte** (*pd.DataFrame*) – Co-variables du jeu de données test

### Renvoie

vecteur contenant les prédictions du modèle sélectionné par cross validation

### Type renvoyé

pred (np.array ou pd.DataFrame)

**general.soumission**(pred, date, name\_pred)

Fonction qui crée une soumission avec le nom de la methode et la sauvgarde dans datasets\_c

### Paramètres

- **pred** (*np.array*) – prediction
- **date** (*str*) – date ex : “0110”
- **name\_pred** (*str*) – nom de la méthode utilisée

**general.train\_eval**(model, X, y, X\_test, y\_test)

Fonction qui entraine un modèle, plotte et renvoie le r2\_score sur le jeu de données test associé à la prédiction

### Paramètres

- **model** (*sklearn.function*) – Modèle à entrainer pour faire la prediction
- **X** (*pd.DataFrame*) – Co-variables du jeu de données train
- **y** (*\_type\_*) – Variable à prédire du jeu de données train
- **X\_test** (*\_type\_*) – Co-variables du jeu de données test
- **y\_test** (*\_type\_*) – Variable à prédire du jeu de données test

### Renvoie

prédiction sur le jeu de données test associé au modèle choisi en input

### Type renvoyé

pred (np.array ou pd.DataFrame)

## CHAPITRE 4

---

### Indices and tables

---

- `genindex`
- `modindex`
- `search`



### d

`dataset_division`, 3

### g

`general`, 5



### B

`build_pred()` (*dans le module general*), 5

### C

`CV_rep()` (*dans le module general*), 5

### D

`dataset_division`  
module, 3

### F

`formal_div()` (*dans le module dataset\_division*), 3

### G

`general`  
module, 5

### M

`module`  
  `dataset_division`, 3  
  `general`, 5

### P

`param_selection()` (*dans le module general*), 5

### S

`soumission()` (*dans le module general*), 6

### T

`train_eval()` (*dans le module general*), 6  
`treatment()` (*dans le module dataset\_division*), 3

### W

`winetype()` (*dans le module dataset\_division*), 3