

# Prédiction des décès dus au Covid19 à l'aide de machine learning

Enzo De Carvalho

numéro d'inscription : 29448

2020-2021

# Sommaire

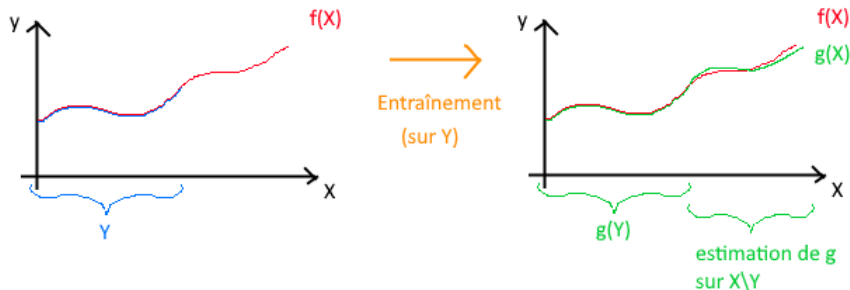
## 1 Première approche : simple regression

- Principe
- Regression Linéaire
- Optimisation d'hyperparamètres
- Résultats avec SVR
  - Cross-validation

## 2 Approche multivariées

- Multiregresseur : 'RegressorChain'
- Réseau neuronal

# Principe de la demarche



↔ le modèle  $\hat{g}$  généralise les données connues  $Y$  fournies.

# Première approche : Regression linéaire

En utilisant **ElasticNet**

Modèle :

$$\hat{g}_{deces}(\omega, t) = \omega_0 + \omega_1 t$$

$t$  les données (le temps)

$\omega = \begin{pmatrix} \omega_0 \\ \omega_1 \end{pmatrix}$  un paramètre à déterminer

# Première approche : Regression linéaire

En utilisant **ElasticNet**

Modèle :

$$\hat{g}_{deces}(\omega, t) = \omega_0 + \omega_1 t$$

$t$  les données (le temps)

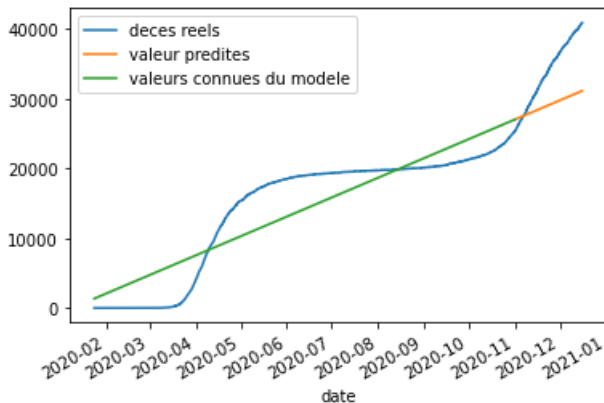
$\omega = \begin{pmatrix} \omega_0 \\ \omega_1 \end{pmatrix}$  un paramètre à déterminer

le modèle **ElasticNet** détermine alors  $\omega$ .

Le résultat dépend des hyperparamètres :  $\alpha$  et  $\rho$

# Première application

Prédictions entre le 2020/11/01 et 2020/12/16.



$$\rho = 0.9$$

$$\alpha = 0.1$$

Figure: Résultats peu satisfaisant... (ici  $l1\_ratio$  est  $\rho$ )

# SVR ; Premier résultat

Modèle **SVR**

Hyperparamètres :

- $C$  le paramètre de régularisation

- $\epsilon$  la taille du tube de "non-pénalité"

- $\gamma$  paramètre du noyau (rbf ici)

# SVR ; Premier résultat

Modèle **SVR**

Hyperparamètres :

$C$  le paramètre de régularisation

$\epsilon$  la taille du tube de "non-pénalité"

$\gamma$  paramètre du noyau (rbf ici)

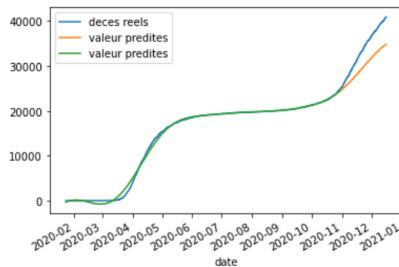
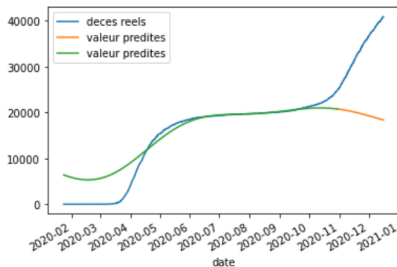
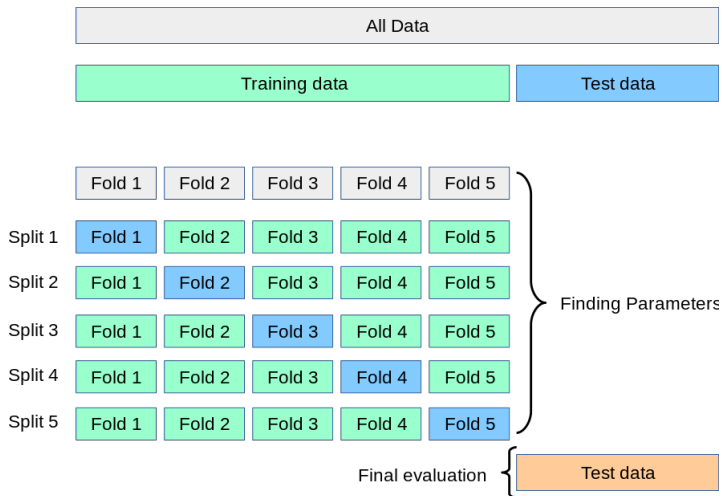


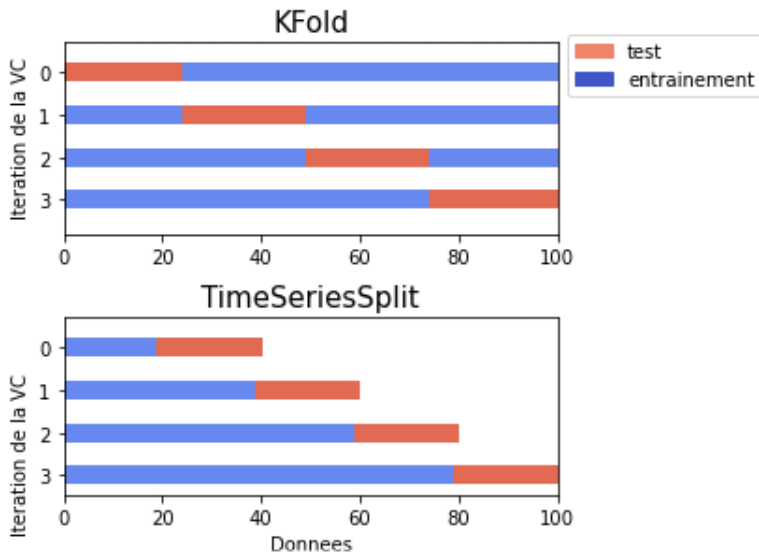
Figure: **SVR** avec  $C = 100$ , puis  $C = 100000$



# Validation croisée

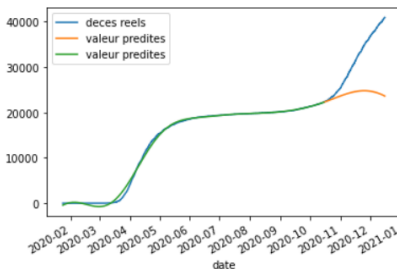


# Stratégie pour la Validation croisée

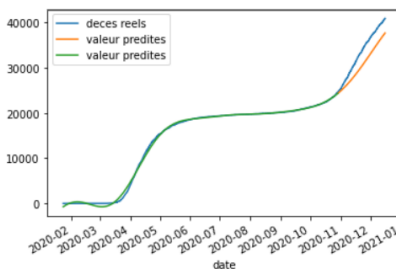


# Application avec SVR

```
params = {'svr__C' : [10**i for i in range(1,8)]}
```



{'svr\_\_C' : 100000}



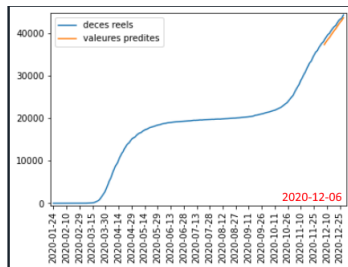
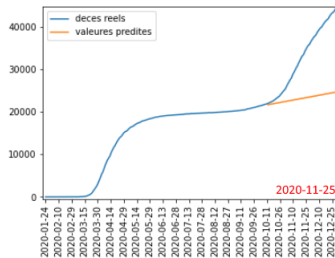
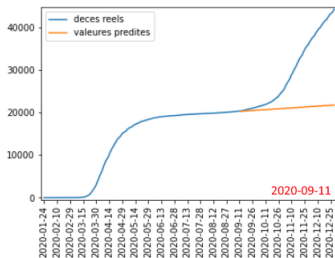
{'svr\_\_C' : 1000000}

Figure: à partir du 2020/10/15, puis du 2020/11/01

⇒ Échec de généralisation.

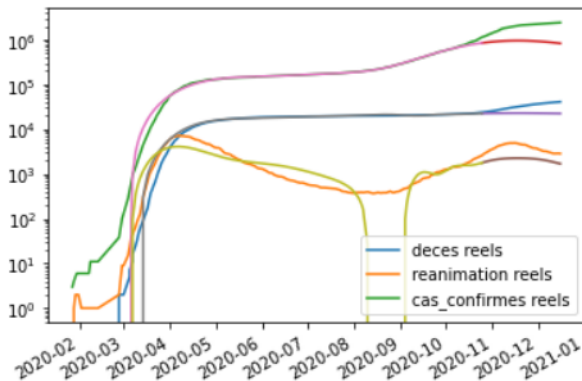
# Prophet

## Approche avec le modèle Prophet de Facebook



# RegressorChain SVR

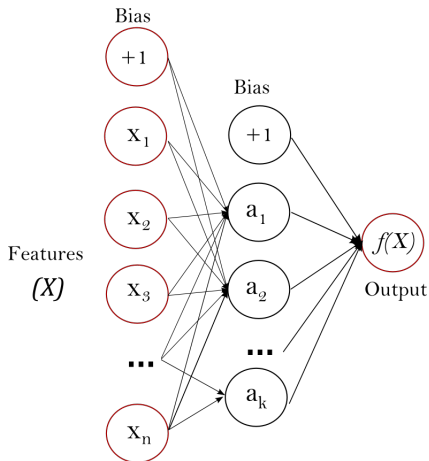
Approche multivariés avec RegressorChain



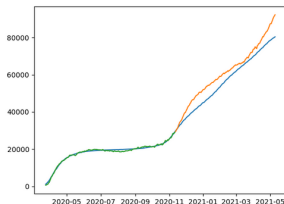
```
{'regressorchain_base_estimator_C': 100000, 'regressorchain_base_estimator_epsilon': 0.001}
```

# Réseau neuronal

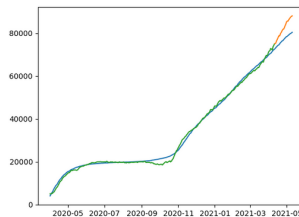
Approche avec un **réseau de neurone**



# Application avec réseau neuronal

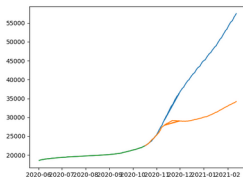


Neural network avec 7 jour de décalage; max\_iter=100k; prédit à partir du 11

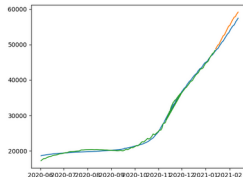


```
{'mlpregressor__max_iter': 90000, 'mlpregressor__n_iter_no_change': 3, 'mlpregressor__tol': 0.0001}
```

NN: décès prédits 30 jours avec les données de 15 jours avant

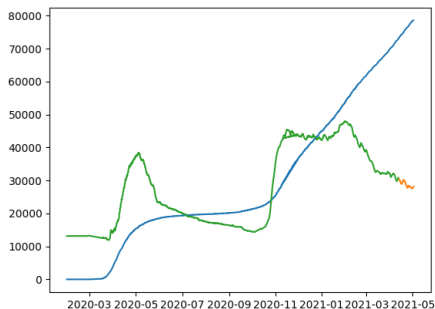


Neural network avec 7 jour de décalage; max\_iter=90k



# Application avec réseau neuronal

	total_cas_confirmes	total_deces_hopital
total_cas_confirmes	1.000000	0.977939
total_deces_hopital	0.977939	1.000000



⇒ échec du modèle sans la courbe des cas confirmés



# fonction d'objectif d'ElasticNet

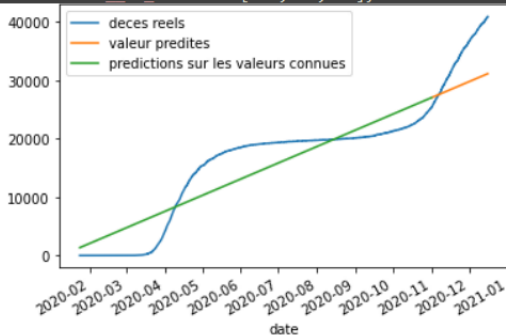
ElasticNet cherche  $\omega$  tel que :

$$\min_{\omega} \frac{1}{2n_{decès}} \|\hat{g}_{decès}(\omega, t) - f(t)\|_2^2 + \alpha \rho |\omega| + \frac{\alpha(1-\rho)}{2} \|\omega\|_2^2$$

$\alpha$  et  $\rho$  les hyperparamètres définissant le modèle,  
 $f$  la courbe réelle des décès.

# Détails sur la regression linéaire

```
params = {'lasso_alpha' : [i/100 for i in range(10,25)]  
          #'svr_degree': [0],  
          #'svr_C': [600000]  
          'elasticnet_alpha' : [i/100 for i in range(10,100)],  
          'elasticnet_l1_ratio' : [0.5,0.8,0.9]}
```



```
deces      622  
dtype: int64  
{'elasticnet_alpha': 0.1, 'elasticnet_l1_ratio': 0.9}
```