

Super titre de TIPE

sous-titre

Alexandrine, Pedro
De Carvalho, Enzo

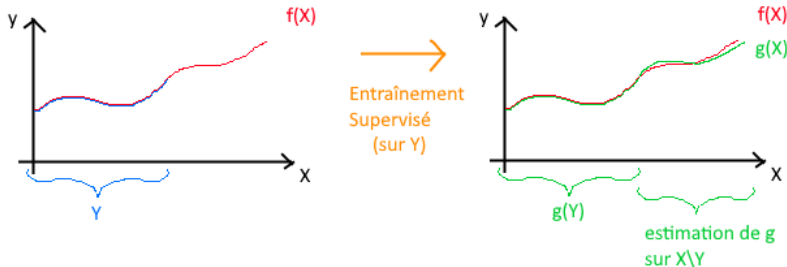
2020-2021

Sommaire

- 1 Première approche : simple regression
 - Principe
 - Optimisation d'hyperparamètres et stratégies
 - Cross-validation
 - Résultats avec SVR
 - Découpages adaptés
- 2 Approche multivariées
 - Multiregresseur : 'RegressorChain'
 - Réseau neuronal

Principe de l'apprentissage supervisé

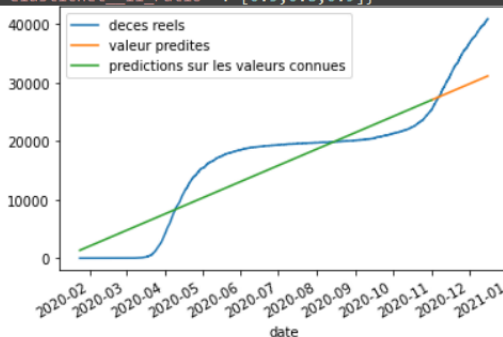
Donné une fonction $f : X \rightarrow y$ (X et y des données corrélées), Un algorithme d'apprentissage supervisé cherche alors à obtenir un modèle $g : X \rightarrow y$ qui estime f à partir d'une partie $Y \subset f(X)$ déjà connue de l'algorithme.



Regression linéaire

ElasticNet ne fait que des droites, c'est pas intéressant.

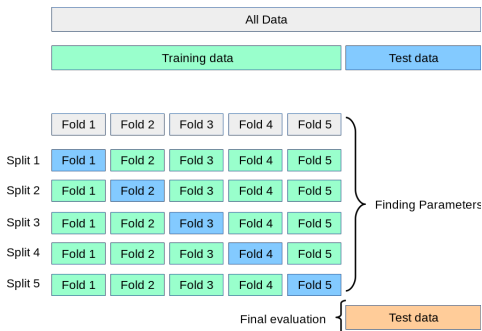
```
params = {'lasso_alpha': [i/100 for i in range(10,25)]  
          #'svr_degree': [0],  
          #'svr_C': [600000]  
          'elasticnet_alpha': [i/100 for i in range(10,100)],  
          'elasticnet_l1_ratio': [0.5,0.8,0.9]}
```



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

Recherche du meilleur paramètre

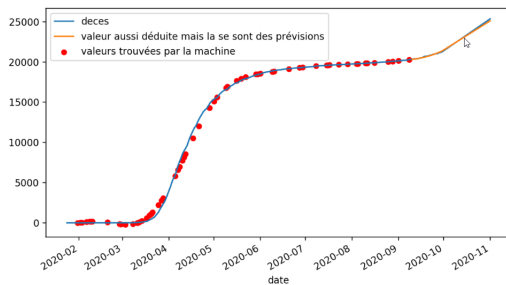
On recherche la meilleur combinaison des paramètres donnés en évaluant successivement chaque modèle pour chacune des combinaisons possibles selon une stratégie de "cross-validation" :



SVR ; premier résultat

Approche à l'aide du modèle SVR (avec le noyau 'rbf') (Support Vector Regressor)

— Modèle SVR (prédit les données entre le 02/12/20 et 16/12/20)



```
{'svr__C': 500000, 'svr__cache_size': 200, 'svr__degree': 0}
0.9980841978694834
```

Figure: Premier résultat avec SVR avec une "mauvaise stratégie" de découpage pour former un modèle prédisant une série temporelle

Découpage adapté pour la CV

Une note moyenne est donnée à un set de paramètres selon un score moyen donné sur chacun des découpages. Pour une évaluation cohérente, il faut un découpage cohérent :

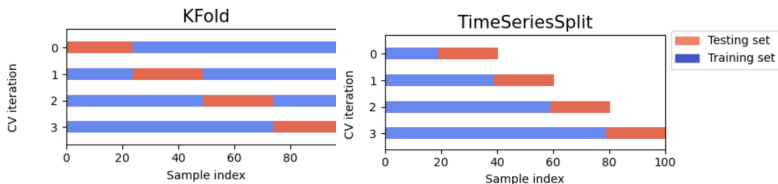


Figure: Comparaison des différents découpages des données employés pour la CV

SVR

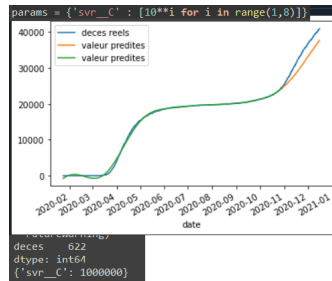
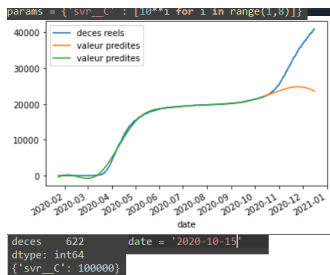


Figure: à gauche la prediction avant le pt d'inflexion, à droite, après.

⇒ Le modèle n'arrive pas à "suivre" sans le point d'inflexion.

fbprophet

Approche avec le module fbprophet (designé pour prédire des séries temporelles)

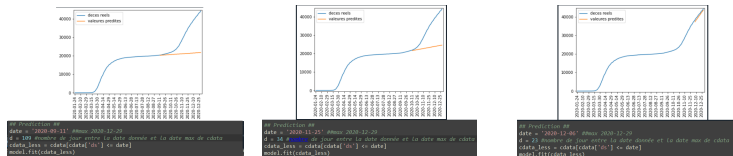


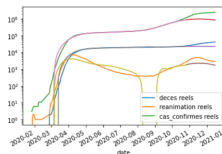
Figure: résultats avec fbprophet (insatisfaisant)

RegressorChain SVR

Approche avec le multiregresseur

(description de scikit : <https://scikit-learn.org/stable/modules/generated/sklearn.multioutput.RegressorChain.html#sklearn.multioutput.RegressorChain>)

"Each model makes a prediction in the order specified by the chain using all of the available features provided to the model plus the predictions of models that are earlier in the chain."



Multiregresseur

```
! [54]: runfile('C:/Users/Anjo/Documents/TYPE/repo/covid_deces_multiregresseur.py', wdir='C:/Users/Anjo/Documents/TYPE/repo')
C:\Users\Anjo\Documents\TYPE\repo\covid_deces_multiregresseur.py:1: DtypeWarning: Columns (17,18)
have mixed types.Specify dtype option on import or set low_memory=False.
### Multi regresseur avec le modele SVR ###
('regressorchain_base_estimator__C': 100000, 'regressorchain_base_estimator__epsilon': 0.001)
```

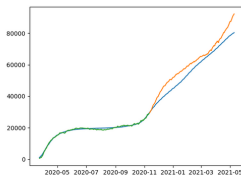
Figures now render in the Plots pane by default. To make them also appear inline in the Console, uncheck "Mute Inline Plotting" under the Plots pane options menu.

Figure: bug qu'on a pas réussi à corriger sous ces conditions

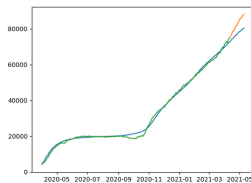
Explications

Là on explique savamment l'art du réseau neuronal

Résultats



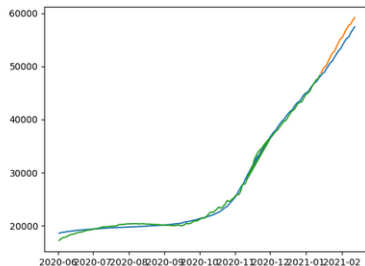
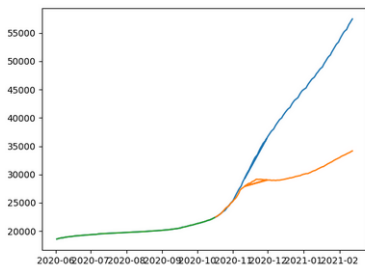
Neural network avec 7 jour de décalage; max_iter=100k; prédit à partir du 11 novembre



{'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

Résultats



Neural network avec 7 jour de décalage; max_iter=90k

Figure: échec du modèle sans la courbe des cas confirmés