

Étude sur le propagation du sars-cov1 avec machine learning

Pedro ALEXANDRINE

numero d'inscription : 41758

2020-2021

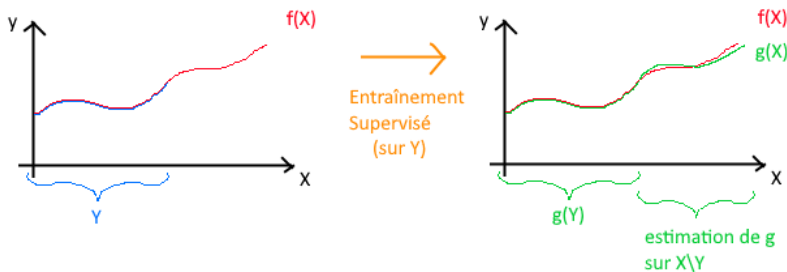
Sommaire

- 1 Première approche : simple regression
 - Principe
 - Validation croisée et hyperparamètres
 - Résultats avec SVR

- 2 Approche multivariées
 - Multiregresseur : 'RegressorChain'
 - Réseau neuronal

Principe de l'apprentissage supervisé

Entraînement donnés X et résultats Y fournis



Principe de l'apprentissage supervisée

Optimisation d'une fonction d'objectif

Forme de modèle de régression linéaire:

$$y(w, x) = w_0 + W_1x_1 + \dots + w_px_p$$

Ici x : données, w : paramètres du modèle

ElasticNet

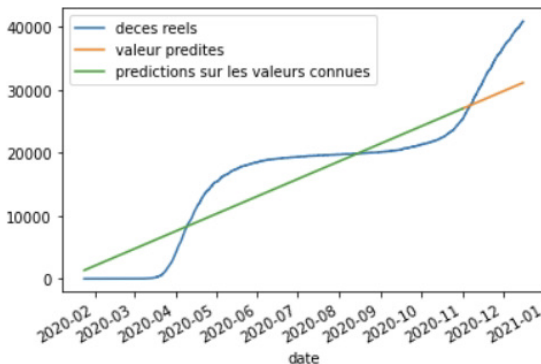
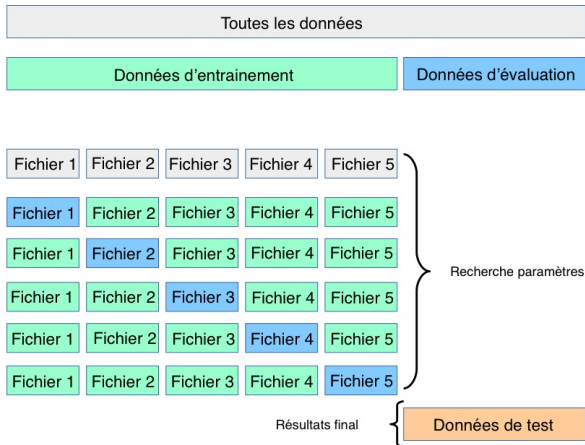


Figure: Exemple avec régression linéaire

Peu intéressant dans notre cas.

Recherche du meilleur paramètre

Principe de la validation croisée:

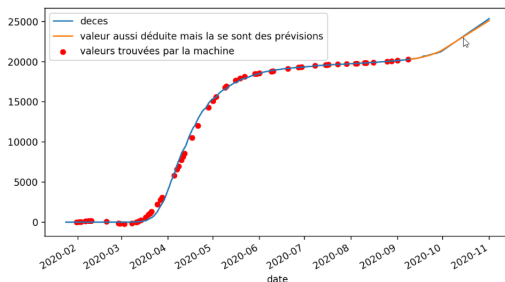


SVR, premier résultat

Approche à l'aide du modèle SVR.

Noyau 'rbf' → ajustement du paramètre C

— 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 et découpage inadapté

Découpage adapté pour la validation croisée

Remise en question de la méthode de découpage

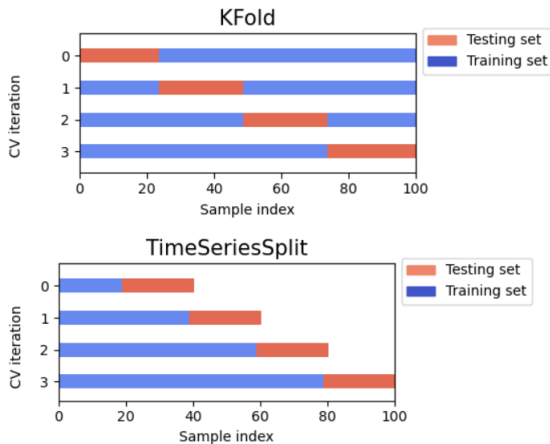


Figure: Comparaison des découpages pour la validation croisée

SVR

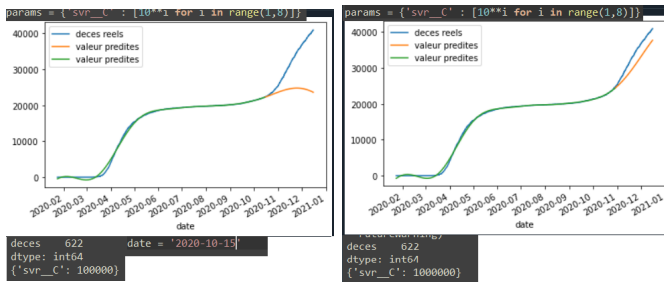


Figure: à gauche la prediction avant le pt d'inflexion, à droite, après.

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

Recherche corrélation

Corrélation pour savoir les données pertinentes

	total_cas_confirmes	patients_hospitalises
total_cas_confirmes	1.000000	0.661548
patients_hospitalises	0.661548	1.000000

	total_cas_confirmes	total_deces_hopital
total_cas_confirmes	1.000000	0.977939
total_deces_hopital	0.977939	1.000000

	total_deces_hospital	patients_hospitalises
total_deces_hospital	1.00000	0.58516
patients_hospitalises	0.58516	1.00000

RegressorChain SVR

Multiregresseur RegressorChain

Corrélation : Cas confirmé → hospitalisé → décès

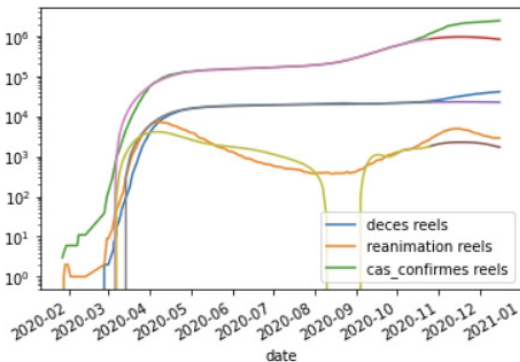
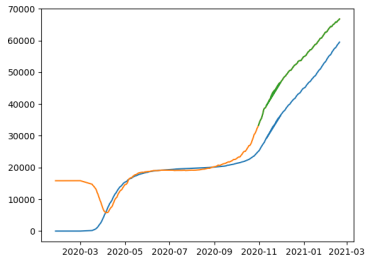
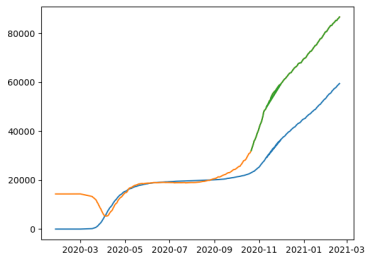


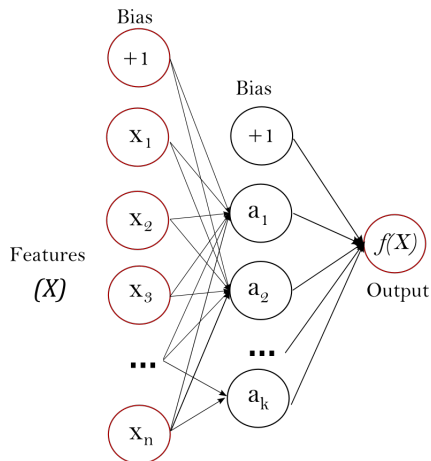
Figure: Résultat avec SVR insatisfaisant

RegressorChain TheilSenRegressor

Changement de régresseur : meilleurs résultats

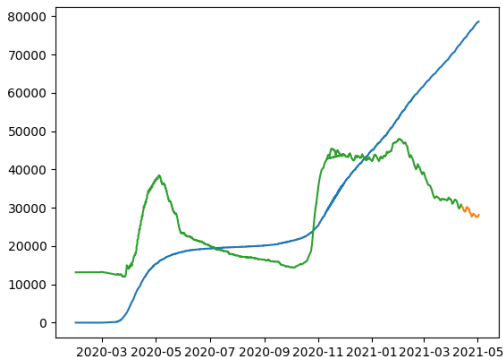


Principe du réseau neuronal



Premier résultat

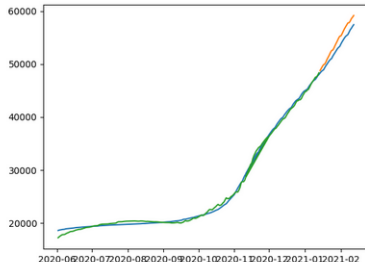
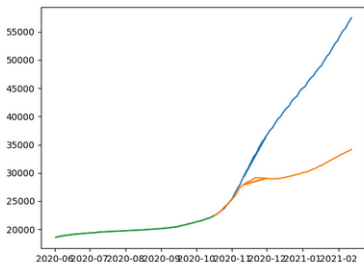
140 000 iteration sans les cas confirmés



Échec du modèle sans les cas confirmés → cohérent avec corrélation

Résultats

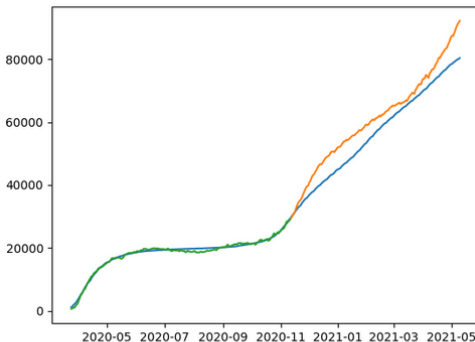
Réseaux neuronaux : Meilleurs paramètres



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

Prédiction durant l'évolution de la situation

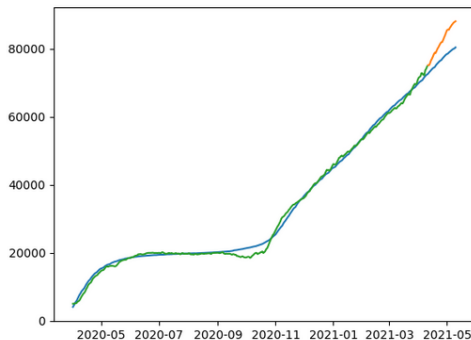
Divergence après quelques jours



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

Augmentation du décalage

Résultats insatisfaisant



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

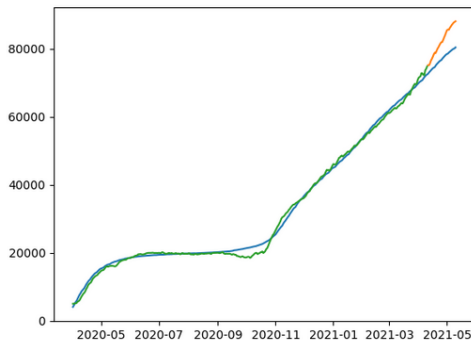
NN: décès prédits 30 jours avec les données de 15 jours avant

Conclusion

Modèle peu fiable hors situation stabilisée ou début d'évolution

Augmentation du décalage

Résultats insatisfaisant



`{'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