

Notice du démonstrateur

Ce démonstrateur comporte 6 onglets. Pour chacun d'entre eux, vous trouverez les explications ci-dessous.

Comment utiliser ce démonstrateur

Page d'accueil. Cliquez sur le bouton de téléchargement afin de pouvoir lire cette notice.

Les SVM : qu'est-ce que c'est?

Si vous n'avez pas de notion en matière de machine learning, nous vous expliquons ici brièvement la méthode des machines à vecteur de support (SVM). Cette courte introduction est inspirée du cours de Christophe Hurlin, que vous pouvez retrouver en intégralité en cliquant sur le lien ci-dessous:

<https://sites.google.com/view/christophe-hurlin/teaching-resources/support-vector-machine>.

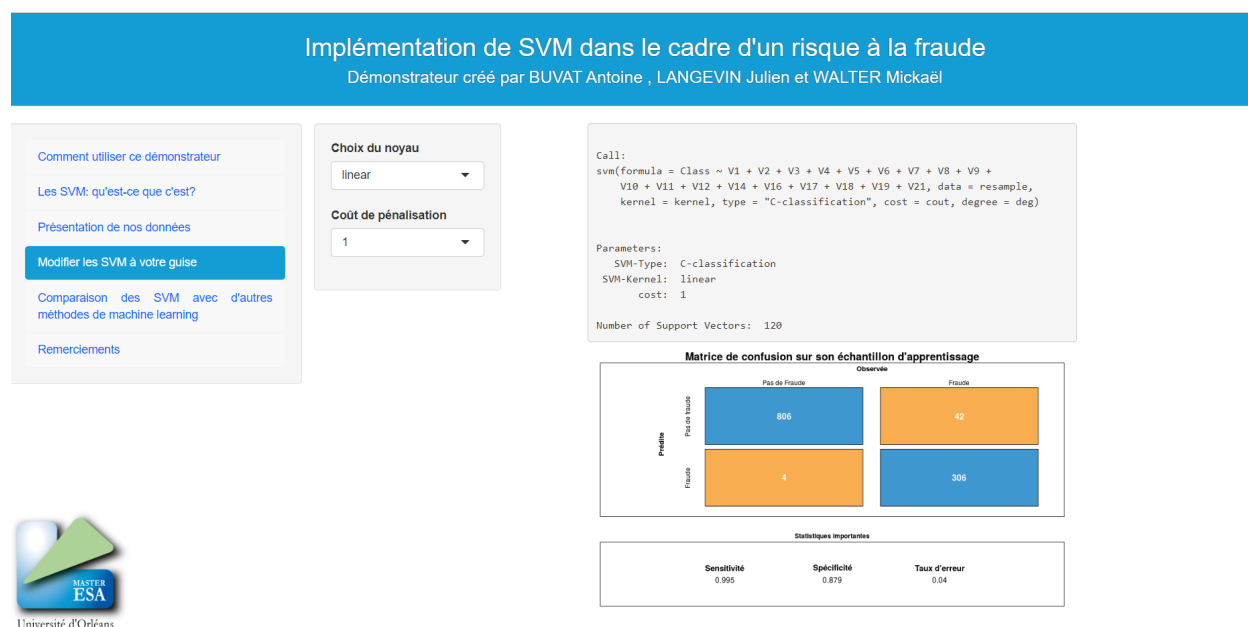
Présentation de nos données

Vous trouverez ici toutes les informations sur les données que nous avons utilisées pour construire ce démonstrateur. Vous pouvez retrouver la base de données sur le site de kaggle, en cliquant sur le lien ci-dessous:

<https://www.kaggle.com/mlg-ulb/creditcardfraud>.

De plus, nous vous expliquons les indices de performances que nous utilisons dans notre analyse.

Interaction avec les SVM



Ici, nous montrons les matrices de confusions sur l'échantillon d'apprentissage et sur l'échantillon test (les observations bien classées sont en bleu, les autres en oranges), la sensibilité, la spécificité et le taux d'erreur qui leurs sont associés, ainsi que la courbe ROC obtenu par la méthode des SVM.

Cette page est interactive. Ainsi, vous pouvez choisir:

- **La fonction noyau** qui permettra de rendre l'échantillon linéairement séparable (linéaire, polynomial, radial ou sigmoïde).
- **Le degré du polynôme** que l'on peut fixer à 3, 4 ou 5 que si nous avons choisi le kernel polynomial.

- **Le coût de pénalisation**, que l'on peut fixer à 1,3, 5 ou 10. Pour rappel, plus ce dernier est important, plus on fait d'erreur, plus il est grand, plus on risque de faire de l'overfitting.

Comparaison des SVM avec les autres méthodes de machine learning

Cet onglet se décompose en quatre parties distinctes:

Recherche des meilleurs SVM

Page d'introduction de l'onglet. On explique brièvement la comparaison des méthodes.

Régression logistique

On compare les résultats des SVM à ceux de la régression logistique.

Random forest

On compare les résultats des SVM à ceux du random forest. Dans cette page, vous pouvez modifier le nombre de prédicteurs qui seront utilisés lors de l'estimation (entre 1 et 18 prédicteurs), ainsi que le nombre d'arbres qui constitueront la forêt (entre 500 et 5000 arbres). Les paramètres initiaux correspondent aux critères optimaux pour cette méthode.

Méthode KNN

On compare les résultats des SVM à ceux de la méthode des plus proches voisins. Dans cette page, vous pourrez modifier le nombre de voisins qui seront utilisés lors de l'estimation de la méthode KNN (allant de 1 à 20 voisins).

Remerciements

Un dernier onglet où nous vous donnons le lien de notre page Github. Sur cette page internet, vous trouverez l'ensemble des codes et des ressources que nous avons utilisées pour la création de ce démonstrateur. Nous citons aussi les personnes qui nous ont été d'une grande importance lors de la création de ce projet.