

Exploiter un modèle d'apprentissage

Compétences visées

- C14. Exploiter un modèle d'apprentissage supervisé ou non supervisé permettant la classification ou la prédiction d'une variable en fonction des données disponibles et des outils sélectionnés
- C15. Améliorer les performances d'un modèle d'apprentissage à l'aide d'une évaluation de la qualité des données et de la technique de modélisation afin de réduire les biais et les anomalies de résultats

Activités et tâches

- Identification du modèle d'apprentissage optimal en fonction du problème à résoudre, des données disponibles et de leurs natures
- Sélection de l'outil (langage, bibliothèque, framework, plateformes)
- Entraînement et exploitation d'un modèle d'apprentissage supervisé à l'aide d'outils préalablement sélectionnés
- Classification ou prédiction d'une variable à partir d'un modèle d'apprentissage supervisé
- Réalisation de divers traitements à l'aide d'un modèle d'apprentissage :
 - langage naturel
 - séries temporelles
 - vision par ordinateur
- Utilisation de l'apprentissage non supervisé pour créer des catégories
- Evaluation de la performance d'un modèle d'apprentissage avec les métriques standards et spécifiques
- Identification des hyper-paramètres du modèle
- Amélioration de données d'apprentissage d'après une analyse des métriques de performance
- Combinaison de plusieurs modèles en un modèle plus performant

Outils envisagés

- Python, R, ...
- Scikit-learn, XgBoost, LightGBM, Keras, numpy, Pandas, opencv...

La régression logistique

- Quel est l'intérêt de la régression logistique ?
- Donner des exemples de problèmes auxquels la régression logistique permet de répondre ?
- Combien y a-t-il de types de régressions logistiques ? Lesquels ?
- Pourquoi le nom de « régression logistique » ?
- Que sont les False Positives et les False Negatives ?
- Qu'appelle-t-on matrice de confusion ? Quel autre nom a-t-elle ?
- Que sont les notions de :
 - performance ou justesse ou accuracy,
 - d'erreur,
 - de précision ou valeur prédictive positive,
 - de sensibilité ou rappel ou recall,
 - de spécificité ?
- Comment interprète-t-on les coefficients de la régression logistique ?
- Que sont les odds ratio (rapport de côtes) ?
- Quel est le lien entre les odds ratio et les coefficients ?
- Que sont les critères ROC et AUC ?

NB : internet est votre meilleur ami

Autre ressource :

<http://www.chups.jussieu.fr/polys/biostats/poly/POLY.Chp.5.3.html>

<https://www.em-consulte.com/rmr/article/143634>

http://cerim.univ-lille2.fr/fileadmin/user_upload/statistiques/michael_genin/Cours/Modelisation/Regression_logistique.pdf