

Rendu Travaux Pratiques 2

Règles de rendu

- Chaque TP donne lieu à un bref compte-rendu portant sur certaines questions posées dans l'énoncé du TP.
 - Le compte-rendu doit être complété à partir du texte de l'énoncé. Les codes R doivent être inclus dans le texte du compte-rendu (menu **Insert**) et commentés avec précision. **Les commentaires compteront pour une part importante dans la note.**
 - Le compte-rendu doit être déposé **sur TEIDE à la fin de la séance de TP**. Les rendus en retard seront fortement pénalisés. Il est
 - Le compte-rendu doit être déposé **sur TEIDE au format HTML uniquement**. Utiliser la fonction **Preview** ou **knitr** du menu de rstudio pour obtenir le document au format souhaité. **Les fichiers "source" (Rmd) ne seront pas acceptés par les correcteurs.**
-

Exercice 1:

- À partir de l'ensemble d'apprentissage, donner les fréquences des variables contenues dans chacune des 9 colonnes du tableau de données `tit_train`. Arrondir les résultats à 2 décimales
- Calculer la probabilité de survie d'un homme (`gender == 1`) ou une femme (`gender == 0`).
- Calculer la probabilité de survie d'un individu sachant sa classe à bord (colonnes 7 à 9 dans le tableau de données `tit_train`).

Exercice 2:

- Calculer les probabilités de survie conditionnelles sachant les variables `gender` et `child` (4 cas).
- Calculer les nombres de prédictions correctes et incorrectes effectuées dans l'ensemble `tit_train` en considérant les fréquences de survie conditionnelles aux variables `gender` et `child`. Représenter les résultats sous forme de matrice de confusion.
- Calculer les nombres de prédictions correctes et incorrectes effectuées dans l'ensemble `tit_train`, le taux de bons classement (*accuracy*) et la perte logloss en considérant les fréquences de survie conditionnelles aux variables `gender`, `child` et `class3` (donner le script complet).
- Pour l'ensemble de test, évaluer les performances du prédicteur fréquentiel s'appuyant sur les caractéristiques `gender`, `child`, `class3` selon les critères *accuracy* et *log-loss* (donner le script complet).
- Pour l'ensemble de test, évaluer les prédictions un prédicteur s'appuyant sur un modèle de réseau de neurones ayant 5 neurones cachés et utilisant le profil complet d'un passager selon les critères *accuracy* et *log-loss* (donner le script complet).
- Quelles prédictions vous semblent les meilleures ? Celles obtenues par les probabilités empiriques ou celles obtenues par le modèle neuronal ? Justifier votre réponse.
- Pour quels individus de l'ensemble test observe-t-on des différences de prédiction entre les approches empiriques et neuronales (donner le script complet) ?