

SY19 – A20
TP 4 (noté) : Régression et classification
Sélection de modèles

Le but de ce TP est de traiter :

- un problème de classification (données `TP4_a20_clas_app.txt`) et
- un problème de régression (données `TP4_a20_reg_app.txt`).

Dans chaque cas, il faudra, en utilisant les méthodes vues en cours, proposer le meilleur prédicteur possible, et une évaluation de la qualité de ce prédicteur (probabilité d'erreur pour le classifieur, espérance de l'erreur quadratique pour le modèle de régression).

Vous êtes libres de la méthodologie à employer. La note du TP dépendra

1. de la variété des méthodes utilisées et de la rigueur de la méthodologie (1/3 des points) ;
2. des performances obtenues sur chacun des deux problèmes (1/3 des points) ;
3. de la qualité du rendu écrit : clarté des explications ; correction du français ou de l'anglais ; qualité des tableaux et des figures ; soin dans la présentation du rapport (1/3 des points).

Vous devrez rendre votre devoir **avant le lundi 2 novembre à minuit** sur Moodle sous forme d'une archive zip (impérativement) contenant exactement *trois* fichiers :

1. Rapport écrit au format pdf réalisé avec un *notebook* RStudio, en français ou en anglais, maximum 8 pages (nom de fichier : **rapport.pdf**)
2. Un fichier **predicteurs.R** contenant deux fonctions, de noms
 - **classifieur**
 - **regresseur**.

Chaque fonction admet comme unique argument un *data frame* contenant les données de test.

3. Un fichier de données R contenant (uniquement) l'environnement nécessaire à l'exécution des deux fonctions ci-dessus (nom de fichier : **env.Rdata**). Vérifiez la taille de ce fichier : les fichiers trop gros ne pourront pas être téléchargés.

Exemple de fichier **predicteurs.R** :

```

classifieur <- function(dataset) {
# Chargement de l'environnement
load("env.Rdata")
# Mon algorithme qui renvoie les prédictions sur le jeu de données
# 'dataset' fourni en argument.
# ...
  return(predictions)
}

regresseur <- function(dataset) {
# Chargement de l'environnement
load("env.Rdata")
# Mon algorithme qui renvoie les prédictions sur le jeu de données
# 'dataset' fourni en argument.
# ...
  return(predictions)
}

```

Remarques :

- Le rapport sera tronqué à 8 pages. Aucune page supplémentaire ne sera pas prise en compte.
- Les fonctions devront s'exécuter automatiquement sans problème. Si ce n'est pas le cas, il ne sera pas tenu compte du résultat.
- Aucun devoir ne sera accepté après la date limite.