

SY19 – A22

TP 4 (noté) : Régression et classification

Sélection de modèles

Le but de ce TP est de traiter :

- un problème de classification (données `TP4_a22_clas_app.txt`) et
- un problème de régression (données `TP4_a22_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. variété des méthodes utilisées et rigueur de la méthodologie (1/3 des points) ;
2. performances obtenues sur chaque problème (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 24 octobre à minuit** sous deux formes :

1. un rapport écrit au format pdf réalisé avec un *notebook* RStudio, en français ou en anglais, maximum 8 pages.
2. un fichier `Rdata` de données R contenant (uniquement) deux fonctions de noms :
 - `prediction_reg`
 - `prediction_cls`.

Chaque fonction admet comme unique argument un *data frame* contenant les données de test et renvoie le résultat de la discrimination/régression.

Vérifiez la taille de ce fichier : les fichiers trop gros ne pourront pas être traités.

Ce fichier de données devra ensuite être chargé sur un site dédié¹ qui calculera les performances de vos algorithmes. Vous êtes limités à 3 essais réussis (3 essais dont les prédictions ont pu être obtenues sans erreur).

Remarques importantes :

1. <http://maggie.gi.utc>

- 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.

Exemple de génération d'un fichier `.Rdata` :

```
# 1. Apprentissage des modèles.

model.reg <- ...
model.cls <- ...

# 2. Création des fonctions de prédiction

prediction_reg <- function(dataset) {
  # Ne pas oublier de charger **à l'intérieur de la fonction** les
# bibliothèques utilisées.
  library(...)

  # Attention à ce que retourne un modèle en prédiction. Par exemple,
# la lda retourne une liste nommée. On sélectionne alors les
# classes.
  predict(clas, test_set)$class
}

prediction_cls <- function(dataset) {
  ...
}

# 3. Sauvegarder sous forme de fichier .Rdata les fonctions
# `prediction_reg`, `prediction_cls`.
# Sauvegarder également les objets utilisés dans ces fonctions
# (`model.reg` et `model.cls` dans l'exemple) !

save(
  "model.reg",
  "model.cls",
  "prediction_reg",
  "prediction_cls",
  file = "env.Rdata"
)
```