

## SY19 – A17

### TP 6 (noté) : Apprentissage à partir de trois jeux de données réelles

Le but de ce TP est de construire des classifieurs aussi performants que possible à partir de trois jeux de données réelles, qui sont brièvement décrits ci-dessous :

**Données expression\_train.txt** : chaque exemple consiste en un vecteur de taille 4200 contenant les niveaux de gris d'une image de taille  $60 \times 70$ , et une étiquette indiquant une expression du visage parmi : joie, surprise, tristesse, dégoût, colère, peur. Pour afficher l'image  $i$ , faire

```
I<-matrix(X.app[i,],60,70)
I1 <- apply(I, 1, rev)
image(t(I1),col=gray(0:255 / 255))
```

**Données characters\_train.txt** : ce jeu de données se compose d'exemples issus de 26 classes correspondant aux 26 lettres de l'alphabet ; chaque vecteur d'apprentissage contient 16 attributs calculés à partir d'une image représentant le caractère correspondant.

**Données parole\_train.txt** : ce jeu de données contient 5 classes correspondant à 5 phonèmes : "sh" comme dans *she*, "dcl" comme dans *dark*, "iy" comme dans la voyelle de *she*, "aa" comme dans la voyelle de *dark*, et "ao" comme dans la première voyelle de *water*. Chaque exemple correspond à la prononciation d'un phonème par un locuteur ; il consiste en un vecteur de 256 attributs extraits du signal de parole et une étiquette de classe.

Votre devoir sera noté sur deux critères :

1. Rigueur de la méthodologie et clarté des explications
2. Performances des classifieurs proposés.

Vous devrez rendre votre travail **avant 5 janvier 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 12 pages (nom de fichier : **rapport.pdf**)

2. Un fichier `classifieurs.R` contenant trois fonctions, de noms
  - `classifieur_expressions`
  - `classifieur_characters`
  - `classifieur_parole`.
 Chaque fonction admet comme unique argument un *data frame* contenant les données de test.
3. Un fichier de données R contenant l'environnement nécessaire à l'exécution des trois fonctions ci-dessus (nom de fichier : `env.Rdata`).

Exemple de fichier `classifieurs.R` :

```
classifieur_expressions <- 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)
}

classifieur_characters <- 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)
}

classifieur_parole <- 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)
}
```

Nous serons cette fois plus stricts que pour le TP noté précédent :

- Le rapport sera tronqué à 12 pages. Toute 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.