**Mode d’emploi**

Les programmations principales sont inclues dans le dossier de fichiers « Travail ».

Il existe trois dossiers de fichiers dans le dossier «Travail », soient « PEbase », « quan&qual » et « breedsdog». Le dossier «PEbase » est le dossier qui traite la prévision de défaut de TPE tandis que les dossiers « quan&qual »et « breedsdog » sont modèle jouet pour étudier la carte Kohonen.

**PEbase :**

Les codes sont enregistrés dans le dossier « code final », et des fonctions des fichiers R sont listés ci-dessous:

**algo-supervisé sélectionné.R :**

Les algorithmes supervisés implémentés avec des variables choisies par des méthodes de la sélection de variables

**algo-supervisé.R :**

Les 4 algorithmes supervisés pour faire la classification

**apprentissage hybride.R :**

L’implémentation de l’apprentissage hybride. Les 4 composantes :1) Regroupement des clients sains 2) le remplacement de valeurs manquantes selon des neurones 3) la constriction des classifieurs locaux 4) mise en œuvre la régression pénalisée

**carte de Kohonen.R :**

L’implémentation de la carte de Kohonen classique sur les données. Les 4 composantes : 1) L’entrainement de la carte 2) évaluation de regroupement 3) 2D dessin de la carte 4) 3D dessin de la carte

**carte de Kohonen supervisée.R:**

L’implémentation de la carte de Kohonen supervisée sur les données. Les 4 composants : 1)Le sous-échantillonnage (undersampling) 2) L’entrainement de la carte supervisée 3) évaluation de regroupement 4) Matrice de confusion

**calibrage.R :**

3 fonctions pour le calibrages : 1) gridsearch.som : la fonction pour calibrer les hyperparamètres de carte de Kohonen par quadrillage 2) gridsearch.ssom : la fonction pour calibrer les hyperparamètres de carte de Kohonen supervisée par quadrillage, il faut faire attention que la proportion de sous-échantillonnage n’est pas comprise 3) gridsearch.hy : la fonction pour calibrer les hyperparamètres de l’apprentissage hybride par quadrillage

**cross validation.R :**

La validation croisée Monte Carlo:1) la régression pénalisée 2) l’apprentissage hybride

**Enveloppe Convexe.R :**

Calculer et visualiser l’enveloppe convexe des clients sains et clients défaillants sur les premières 2 composantes principales de l’ACP.

**feature selection.R :**

Les 7 méthodes de la sélection de variables qui sont utilisées dans le rapport.

**traitement de donnée.R :**

il fait trois choses : 1) trouver les variables qualitatives selon ses noms et réorganiser la base 2) discrétiser les variables qualitatives 3)diviser la base en base d’apprentissage et base de test

**tsne.R :**

Visualiser les points d’observations dans le plan de 2D en utilisant l’algorithme t-distributed stochastic embedding.

Il existe aussi plusieurs fichiers R à introduire au besoin:

**dist.R :**

dist.weight : distance euclidienne pondérée

distna.weight : distance euclidienne pondérée qui permet des valeurs manquantes

**evaluation.R :**

2 méthodes d’évaluation de regroupement :1) Pureté 2) NMI (normalized mutual information)

**gini plot.R :**

Giniplot : la fonction pour dessiner la figure de Gini des classifieur

**Kohonenprep.R :**

2 fonctions :1) vois.constr pour définir le voisinage sur la carte de Kohonen selon la distance Manhattan 2) para.expo pour définir le comportement de décroissance exponentielle des hyperparamètres

**Kohonennet.R :**

Kohonenqualigo : la fonction d’entrainement de neurones

Kohonenqualiclass : la fonction d’attribuer les observations aux neurones selon la distance euclidienne

**Kohonennet\_weight.R :**

kohonenqualigo.weight : la version pondérée de la fonction Kohonenqualigo

Kohonenqualiclass : la fonction d’attribuer les observations aux neurones selon la distance euclidienne pondérée

**breedsDogs :**

**BreedsDogs.R :**

La programme pour regrouper les genres des chiens à partir de la carte de Kohonen.

**quan&qual :**

**iris.R:**

L’implémentation de la carte de Kohonen sur la base de donnée « iris »

**Cleve.R :**

L’implémentation de la carte de Kohonen sur la base de donnée « cleveland data »

**TWmesure.R :**

L’implémentation d’une carte de Kohonen pondérée sur la base de donnée « cleveland data »  en utilisant la métrique de Shih(2010).