

I. Prétraitement des données

Voir fichier : projet.m

II. Classification des données par la méthode du plus proche voisin (1-NN)

Voir fichier classification.m, tauxClassification.m et tauxClassificationParClasse.m

Avec 100% de Xa et 100% de Xt, on obtient un taux de classification général de **99,869%**.

```
>> tauxClassification(y_pred, yt)
Performannce : 99.869 %
```

```
ans =
```

```
99.8690
```

Classe	Performance
1	99,9448 %
2	99,9793 %
3	99,9655 %
4	100 %
5	99,9931 %
6	99,9931 %
7	99,9931 %

Le temps de calcul de la classification, pour 100% des données de test et d'apprentissage, est en moyenne de **173s**.

III. Classification des données par la méthode de distance minimum

Voir fichier classification2.m

Taux de réussite : 70,0483 %

Temps d'exécution : 8s (21 fois plus rapide que l'algorithme précédent).

Classe	Performance (8s)	Performance algo. Précédent (173s)
1	83,6828 %	99,9448 %
2	99,9103 %	99,9793 %
3	99,7310 %	99,9655 %
4	86,9103 %	100 %
5	99,8207 %	99,9931 %
6	100 %	99,9931 %
7	99,9931 %	99,9931 %
Toute classes	70,0483 %	99,8690 %

IV. Echantillonnage *bootstrap* des données

Voir fichier `echantillonnageBootstrapSansRemise.m` et `echantillonnageBootstrapAvecRemise.m`

Sans remise et avec remise : 86,0828% de réussite

Temps d'exécution moyen : 10s

Classe	Sans remise et avec remise
1	93,4276 %
2	99,9103 %
3	99,7517 %
4	93,1379 %
5	99,8966 %
6	99,9724%
7	99,9862 %

Lorsque nous utilisons `rng('default')`, les performances sont rigoureusement les mêmes peu importe la méthode de tirage. Si nous n'utilisons pas cette ligne, alors cela dépend des fois. Parfois avec tirage devient meilleur, parfois sans tirage devient meilleur.

V. Sélection d'attributs

Voir fichier `selection_dattributs.m`

On a affiché la performance sans la colonne 1, puis sans la colonne 2...

```
Performannce : 99.3379 %  
99.3241 99.1172 99.2552 99.7655 99.4207 99.3517 99.3103 99.3931 99.3379  
Le paramètre 2 est le plus important
```

Activer Win
Appuyer sur...

En retirant la colonne 2, on remarque la plus importante baisse de performance. **Le paramètre 2 est donc le plus important.**

VI. Classification binaire

Q°1 : Voir fichier `binarisation.m`

Q°2 : Voir fichier `classification_binaire.m`

Q°3

Voir fichier classification_binaire.m

Performance générale : 99,8966 %

Temps d'exécution moyen : 174s

Classe	Performance
0	93,9517 %
1	99,9448 %

Q°3 et Q°4

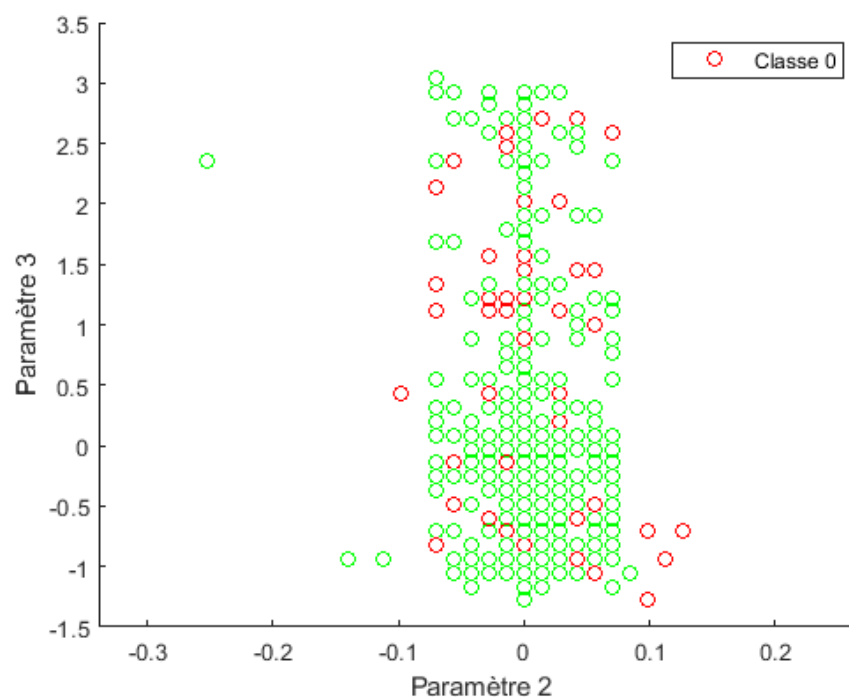
Les deux paramètres les plus importants pour une classification binaire sont les paramètres en colonne 2 et 3

Temps d'exécution : 240s

Paramètres	Performance sans la classe
1	93,4759 %
2	99,2000 %
3	99,3931 %
4	93,8621 %
5	99,5034 %
6	99,4345%
7	99,4207 %
8	99,4759 %
9	99,4483 %

Q°5

Voir fichier affichage.m



VI. Classification des données par la méthode des K plus proches voisins (K-NN)

Voir fichier knn.m

Pour un temps d'exécution 3 fois plus longs, les performances sont de 71%, ce qui semble contradictoire.