

Apprentissage : TP

Comparaison PMC/SVM

1 Objectif

Le but de ce TP est de comparer SVM et PMC sur un même jeu de données. Pour la partie PMC, vous utiliserez une toolbox répandue, Netlab.

1.1 Outils

- Le code sera en Matlab.
- Les réseaux de neurones utilisés sont ceux de la boîte à outils Netlab (disponible gratuitement en ligne).
- Les données sont issues de la base de données USPS.

1.2 Etapes

- Préparer les bases d'apprentissage et de test.
- Etudier l'utilisation du code pour le perceptron multi-couches (PMC).
- Tester le PMC sur les données synthétiques afin de maîtriser les différents paramètres.
- Mettre en œuvre le PMC sur les images issues de USPS et comparer les performances avec celles des SVM.

2 Préparation des données

Nous allons travailler sur 2 bases différentes. La première consiste en un problème de classification binaire en 2D non séparable. La seconde est un problème de classification binaire extrait de la base USPS, une des classes contre une des autres (cf. TP précédent).

Pour chaque base, il faut générer la base d'apprentissage avec ses étiquettes ainsi que la base de test avec ses étiquettes.

1. Base 1 : les données sont générées à partir de deux distributions gaussiennes de centre et de variance différents, de façon à ce que les classes se chevauchent partiellement (figure 1)
2. Base 2 : cf. TP précédent.

3 Le perceptron multi-couche (PMC)

La boîte à outils netlab contient une implémentation d'un PMC. L'aide (*help mpl*) et les fichiers d'exemple *demmpl1.m* et *demmpl2.m* (dans le répertoire de la toolbox) vous permettront d'apprendre l'utilisation de cette boîte à outils.

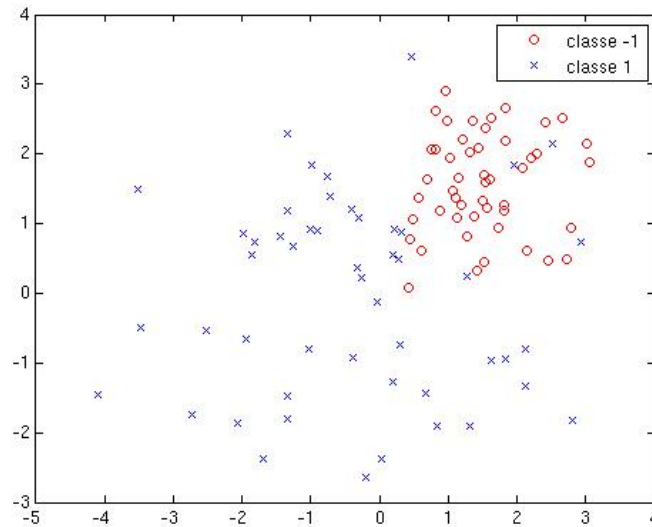


FIGURE 1 – 2 classes gaussiennes

Remarque : les sorties d'un PMC sont continues entre 0 et 1, il faut penser à les convertir en $[-1, +1]$...

1. Utiliser la base 1 avec le PMC. Etudiez l'impact des paramètres sur la qualité de l'apprentissage (nombre de neurones cachés...).
2. Appliquer la même procédure sur la base 2 (attention, il faudra plus de neurones cachés et plus d'itérations, cela peut devenir long). Votre objectif est de battre le SVM en termes de taux d'erreur (cf. section suivante : y arriverez vous ??).

Commandes utiles

```
net = mlp(nin, nhidden, nout, 'logistic', alpha);
[net] = netopt(net, options, trainvec, trainlab, 'quasinew');
ypred = mlpfwd(net, testvec);
ypred(find(ypred>0.5)) = 1;
ypred(find(ypred<=0.5)) = -1;
```

4 Comparaison SVM/PMC

Reprendre vos précédents résultats (SVM) et les résultats de ce TP pour comparer les performances des deux algorithmes sur la base USPS. Le compte rendu doit contenir une description des essais réalisés, et la motivation du choix final. Les méthodes d'évaluations à utiliser sont (au moins) la validation croisée et le taux de mauvaise classification. Les illustrations, résultats numériques, choix de découpage de la base d'apprentissage, interprétations etc. doivent figurer dans le rapport.