

Classement de caractère manuscrits

La reconnaissance de caractères manuscrits par un système automatisé est un problème aux multiples applications: reconnaissance des codes postales, création de petit système mobile de saisie de texte (par exemple, pour les ordinateurs de poche), numérisation de documents manuscrits, etc. Les méthodes les plus performantes pour reconnaître un caractère manuscrit sont basées sur des méthodes d'apprentissage statistique: leur principe commun est de fonder leur prédiction sur la comparaison de l'image du caractère manuscrit à classer à d'autres images de caractères manuscrits pour lesquels la nature du caractère est connue. Par exemple, si une image arrive et qu'elle est très similaire à l'image de nombreux '1' de notre base d'apprentissage, l'algorithme classera l'image dans la catégorie '1'.

Nous allons examiner l'algorithme Régression Logistique sur une base de données d'images de chiffres manuscrits au format 20 pixels par 20 pixels où chaque pixel est représenté par un niveau de gris allant de 0 à 255 (i.e. un chiffre manuscrit est un vecteur de $\{0, \dots, 255\}^{20 \times 20}$). L'objectif est seulement de se familiariser avec Régression Logistique. Le code est déjà écrit. Suivez les instructions suivantes.

- Téléchargez le support de TP "TP7.zip"
- Téléchargez la base de données "ex3data1.mat"
- Exécutez "LR.sci" en scilab et suivez interactivement le code, pour vous familiariser avec l'algorithme Régression Logistique.

À rendre

Un rapport qui explique Régression Logistique et ses étapes principales.
(Envoyer par e-mail à azadeh.khaleghi@inria.fr)

La date limite de rendu est le mardi 12 mars 2013 à 23:59