

TP6 – SVM<sup>light</sup>

{Francois.Denis, Yann.Esposito}@lif.univ-mrs.fr  
26 janvier 2005

- ☞ Ce TP sera dédié à la prise en main de SVM<sup>light</sup>. En particulier une partie du travail sera consacrée à la vérification du comportement des SVM et de filtres de prétraitements.

## 0 Récupération de SVM<sup>light</sup> et des données

Récupérez le binaire de SVM<sup>light</sup> à l'adresse suivante : <http://svmlight.joachims.org/>

Lisez les informations sur la page, exécutez le premier exemple proposé.

Cette fois-ci nous allons travailler avec des photographies de visage. Les données sont disponibles sur mon site :

<http://www.lif.univ-mrs.fr/~esposito/enseignement.php>

Téléchargez et décompressez l'archive. Celle-ci contient toutes les images au format `pgm` ainsi que des scripts de générations d'échantillons pour `svmlight`. Ces scripts utilisent un programme écrit en C++ qui transforme un fichier `pgm` en une ligne du fichier échantillon.

Toujours sur mon site vous pouvez voir les visage. L'ensemble d'entraînement est constitué des cinq premiers visage de chaque personne. Les tests seront à effectués avec les cinq autres photos.

## 1 Utilisation de filtres

**1.1.** Expliquez comment classer dans  $n$  classes avec des SVM.

**(a)** Dans un premier temps, essayez d'apprendre un modèle qui reconnaît seulement la première personne. Testez le taux de réussite, est-ce bon ? (Les jeux de données sont `train-1` et `test-1`)

**(b)** Proposez un post-traitement facile à mettre en œuvre qui améliore la situation.

**1.2.** L'apprentissage en considérant toutes les données brutes est assez long.

**(a)** Donnez une méthode qui permette à la fois de diminuer le nombre d'attributs et qui n'est pourtant pas trop destructrice. (indice : en augmentant le nombre d'exemples).

**(b)** Utilisez la méthode précédente pour aussi diminuer la taille des jeux de données.

**(c)** Toujours dans l'archive, vous pouvez trouver tout ce qui est nécessaire à la transformation d'une image en fichier de données. Modifiez celui-ci pour effectuer des filtres qui pourront améliorer l'apprentissage. Notamment, modifiez le pour utiliser la méthode vue précédemment. Vous avez aussi le droit d'utiliser des outils comme `convert`, `Gimp`, `xv` ou tout autre outil de manipulation d'image.

## 2 Apprentissage proprement dit

Le taux d'erreur de méthodes déjà non naïve est de 13%. Avec des méthodes poussées, le taux d'erreur arrive à 5%. Il est remarquable que ce ne soit pas l'algorithme d'apprentissage qui soit critique dans l'apprentissage, mais les filtres que l'on propose sur les données. Ainsi, il est tout à fait possible, même avec des idées qui peuvent sembler naïves d'approcher des taux de réussites proches de ceux cités précédemment.

**2.3.** Décrivez une méthode de reconnaissance de visage en utilisant SVM<sup>light</sup> et implémentez-la. Bien-sûr vous pouvez utiliser les scripts que j'ai fourni sur ma page.

Vous m'enverrez sur mon adresse email un résumé de la méthode que vous avez choisit, ainsi que l'implémentation que vous avez fait. Donnez aussi les justifications de vos choix. Vous joindrez à tout ceci, les résultat (taux de reconnaissance).

Il faut absolument terminer avant la fin de ce TP, le prochain ne sera pas consacré aux SVM.

Bonne chance.

