

TD 10 : INÉGALITÉS DE MARGE ET SVMS

COURS D'APPRENTISSAGE, ECOLE NORMALE SUPÉRIEURE, PRINTEMPS 2013

Rémi Lajugie
remi.lajugie@ens.fr

RÉSUMÉ. Ce dernier TP doit vous faire réviser quelques unes des notions importantes vues en cours ainsi que vous faire étudier les bornes et théorèmes vues lors du dernier cours.

1. EXERCICE : AUTOUR DU COURS ET DE QUELQUES FONCTIONS DE PERTE

On rappelle la forme générale des problèmes que l'on cherche à résoudre dans le cadre de ce cours d'apprentissage statistique :

$$(1) \quad \min_{w \in \mathbb{R}^d} \frac{1}{n} \sum_i^n \ell(y, w^T \phi(x_i)) + \lambda \Omega(w)$$

où ℓ est une fonction de perte et Ω une fonction de régularisation.

1) a) Rappelez brièvement quelles sont les pertes associées à la régression linéaire et aux SVMS. Pourquoi n'optimise t'on pas directement l'erreur de classification ?

b) Quelle fonction de régularisation a-t-on utilisé le plus souvent ?

2) Justifiez que pour tout problème de minimisation de la fonction de perte convexe non régularisée soumise à la contrainte $\|x\|^2 \leq t$, il y a un problème régularisé équivalent pour un certain λ^* .

3) On considère un n-échantillon d'apprentissage i.i.d $(X_1, Y_1) \dots (X_n, Y_n)$ où les Y_i sont binaires. On considère une fonction de perte ℓ , et on introduit $m(\theta) = \mathbb{E}[\ell(Y_i, w^T \phi(X_i))]$. En suivant les notations du cours, on introduit $\theta^* = \operatorname{argmin}_{B_d} m(\theta)$ et le minimiseur de la perte empirique $\hat{\theta} = \operatorname{argmin}_{B_d} \frac{1}{n} \sum \ell(Y_i, w^T \phi(X_i))$. Rappelez les résultats du cours reliant $m(\theta^*)$ et $m(\hat{\theta})$ (hypothèses et vitesses de convergence).

4) Les hypothèses sont elles satisfaites pour la perte quadratique? Pour celle des SVM?

5) On introduit la fonction de perte de Huber par :

$$(2) \quad \begin{cases} \ell(u) = \frac{1}{2}u^2, & |u| \leq r \\ \ell(u) = \frac{1}{2}r|u| - r^2/2, & \text{sinon} \end{cases}$$

a) Représentez cette fonction graphiquement avec l'aide de MATLAB/Octave.

- b) Que vaut u dans le cas de la classification binaire supervisée ?
- c) Vérifiez que les hypothèses des théorèmes énoncés précédemment sont vérifiées.

2. IMPLÉMENTATION : ALGORITHMES DE SVMs

- 6) Rappelez quel est la fonction de perte associée aux SVMs. (Support Vector Machine en anglais ou Séparateurs à Vaste Marge en Français).
- 7) Récupérez le package LIBSVM (disponible sur <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>).
- 8) Pour vous familiariser avec son usage, lisez le readme pour comprendre quels arguments il lui faut en entrée (il suffit en résumé de rentrer les labels d'entraînement ainsi que la matrice de design) et ce qu'il peut vous renvoyer.
- 9) Générez des données en dimension 2 suivant deux nuages gaussiens de moyennes distinctes. Ecrivez votre code de sorte que vous puissiez faire varier la marge globale entre les deux nuages facilement.
- 10) Faites une séparation en ensemble d'entraînement et de test.
- 11) Entraînez un SVM sur l'ensemble d'entraînement en réglant arbitrairement le paramètre de régularisation.
- 12) Visualisez maintenant les vecteurs de support associés au classifieur que vous avez appris. Rappelez comment on peut les interpréter en termes de variables duales.
- 13) Implémentez désormais la sélection de λ par validation croisée. (faites le vous même, sans utiliser l'option native de LibSVM...)