

Optimisation du Primal pour les SVM

Trinh-Minh-Tri Do*, Thierry Artières*

*LIP6, Université Pierre et Marie Curie
104 avenue du Président Kennedy, Paris, France
{Trinh-Minh-Tri.Do, Thierry.Artieres}@lip6.fr

Résumé. L'apprentissage de SVM par optimisation directe du primal est très étudié depuis quelques temps car il ouvre de nouvelles perspectives notamment pour le traitement de données structurées. Nous proposons un nouvel algorithme de ce type qui combine de façon originale un certain nombre de techniques et idées comme la méthode du sous-gradient, l'optimisation de fonctions continues non partout différentiables, et une heuristique de shrinking.

1 Introduction

Les Machines à Vecteurs de Support (SVM) sont une méthode très populaire d'apprentissage supervisé pour la classification et la régression. Dans sa forme la plus simple pour la classification bi-classes, cette méthode est basée sur un classificateur linéaire séparant deux ensembles de points par un hyperplan. L'idée originale est de trouver un hyperplan séparant "au mieux" les points par la maximisation de la marge entre l'hyperplan séparateur et les points dans la base d'apprentissage. Cette formulation conduit à un problème d'optimisation d'une fonction convexe sous des contraintes linéaires. Récemment des extensions de cette technique de base et de l'approche de maximisation de la marge ont été proposées pour le traitement de données structurées comme les séquences, les arbres etc (Tsochantaridis et al., 2004).

La méthode originale de Vladimir Vapnik pour résoudre le problème d'optimisation avec contraintes des SVMs consiste à introduire des multiplicateurs de Lagrange pour chaque contrainte, et d'optimiser le problème dual équivalent. Cet algorithme est coûteux en temps et en mémoire. Par exemple, l'espace mémoire nécessaire (la matrice noyau est de taille N au carré, si N est le nombre d'exemples). Ces caractéristiques de complexité rendent difficile l'emploi de machines à vecteurs support et plus généralement de méthodes de maximisation de la marge dans certaines situations, lorsque l'on traite des données structurées ou bien lorsque l'on dispose de très grandes quantités de données d'apprentissage. Plusieurs voies ont été suivies pour dépasser les problèmes posés par l'optimisation dans ce cadre.

Certains travaux ont porté sur l'optimisation efficace du dual, par le contrôle du nombre de contraintes actives (Joachims, 2006), ou par la décomposition du problème d'apprentissage (Osuna et al., 1997). Dans ce dernier cas, l'algorithme SMO ou SVMLight par exemple, on ne s'intéresse à une itération donnée qu'à un nombre limité de variables actives.

Des travaux plus récents ont porté sur l'optimisation directe de la forme primale par l'usage de la fonction $\text{hinge}(z) = \max(0, z)$. Cela permet de se ramener à un problème d'optimisation sans contraintes où la fonction objectif est convexe. La difficulté de ces dernières approches