Projet Optimisation
Support Vector Machine

K. Kamtue & Cl. Réda

ENS Cachan

December 19, 2016

Implémentation d'un SVM

Objectif : Faire de l'apprentissage supervisé

- Appliqué à la classification binaire

[scale=0.3]images/voronoi.png

- Recherche d'une frontière linéaire f vérifiant (condition 1) :

$$\forall i, y_i = -1 \Rightarrow f(x_i) \leq -1$$
$$\forall i, y_i = 1 \Rightarrow f(x_i) \geq 1$$
$$\Leftrightarrow \forall i, y_i \times f(x_i) \geq 1$$

Recherche du problème d'optimisation

Naïvement

$$max_w \ \gamma = \frac{2}{\|w\|} \text{ avec } (1)$$
$$\Leftrightarrow min_w \ \|w\| \text{ avec } (1)$$
$$\Leftrightarrow min_w \ \frac{1}{2} \times \|w\|^2 \text{ avec } (1)$$

Amélioration

En rendant le problème toujours faisable

(P) $min_w \ \frac{1}{2} \times \|w\| + C \times \sum_{i \leq m} z_i$
avec $\forall i, z_i \geq 0$
$\forall i, y_i \times (\omega^T x_i) \geq 1 - z_i$

Résolution du problème d'optimisation

- Utilisation de la méthode de Newton :

Rappel

$$\omega_{n+1} \leftarrow \omega_n + size \times \nabla^2 obj(\omega_n)^{-1} \nabla obj(\omega_n)$$

(ici, en cherchant *size* par *backtracking line search*)

- Rendre le problème indépendant de la *dimension*, ce qui permet d'appliquer cette méthode pour des problèmes à dimension de taille modérée (mais le problème sera dépendant du nombre d'échantillons !) en résolvant le *problème dual*
- Utiliser la méthode de la barrière logarithmique pour s'affranchir des contraintes d'inégalité

Le problème dual

Après calcul du lagrangien et minimisation par rapport à ω et z :

Problème dual

$$max_{\lambda \in \mathbb{R}^{+m}} - \frac{1}{2} \| \sum_i \lambda_i y_i x_i \|_2^2 + \mathbf{1}^T \lambda$$
$$\text{avec } \forall i, 0 \leq \lambda_i \leq C \text{ si } z_i > 0$$

Obtenir la solution du primal à partir de celle du dual

$$\omega^* = \sum_i \lambda_i^* y_i x_i$$

Rendre le problème indépendant de la dimension

Utilisation de l'*astuce du noyau* :

Problème dual

Soit $K = X^T X$. Alors :

$$\max -\frac{1}{2}\lambda^T diag(y) K diag(y)\lambda + \mathbf{1}^T\lambda$$
$$\text{avec } \forall i, 0 \leq \lambda_i \leq C$$

Supprimer les contraintes d'inégalité

Utilisation de la *méthode de la barrière logarithmique* :

Fonction barrière

$$\Phi(\lambda) = \sum_i(-log(C-\lambda_i) - log(\lambda_i)) = \sum_i log(\frac{1}{(C-\lambda_i)\lambda_i}) = -\sum_i log((C-\lambda_i)\lambda_i)$$

Problème d'optimisation final

$$max\ -\frac{1}{2}\lambda^T diag(y) K diag(y)\lambda + \mathbf{1}^T\lambda + \Phi(\lambda)$$

Extensions

- Validation croisée pour le choix de la meilleure valeur de C
-