Projet Optimisation
Support Vector Machine

K. Kamtue & Cl. Réda

ENS Cachan

December 29, 2016

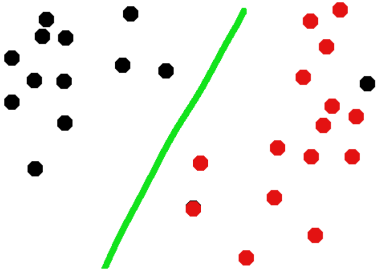
Sujet

Support Machine Vector

Objectif : Faire de l'apprentissage supervisé

- Appliqué à la classification binaire

Figure: Exemple avec deux classes (rouge et noire)



Le problème d'optimisation

Recherche du problème d'optimisation

Naïvement

γ est la distance entre les droites $f(x) = 1$ et $f(x) = -1$.

$$max_w \ \gamma = \frac{2}{\|w\|}$$
$$\text{avec } \forall i, y_i f(x_i) \geq 1$$

$$\Leftrightarrow min_w \ \|w\|$$
$$\text{avec } \forall i, y_i f(x_i) \geq 1$$

$$\Leftrightarrow min_w \ \frac{1}{2}\|w\|^2$$
$$\text{avec } \forall i, y_i f(x_i) \geq 1$$

Le problème d'optimisation

Amélioration

En rendant le problème toujours faisable

Pénaliser les erreurs de classification avec les $(z_i)_i$ et C :

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

Implémentation

Résolution du problème d'optimisation

- Utilisation de la méthode de Newton :

Rappel : Mise à jour du vecteur x avec la méthode de Newton

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

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

- Rendre le problème indépendant de la *dimension* (**dépendant du nombre d'échantillons** !) en résolvant le *problème dual*;
- Utiliser la méthode de la barrière logarithmique.

Implémentation

Le problème dual

Après calcul du lagrangien, minimisation en ω et z (λ multiplicateur de Lagrange) :

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

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

Implémentation

Rendre le problème indépendant de la dimension

Utilisation de l'*astuce du noyau* :

Problème dual

Soit $K = X^T X$ (noyau). Alors :

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

Implémentation

Supprimer les contraintes d'inégalité

Utilisation de la *méthode de la barrière logarithmique* :

Fonction barrière pour éliminer les contraintes d'inégalité

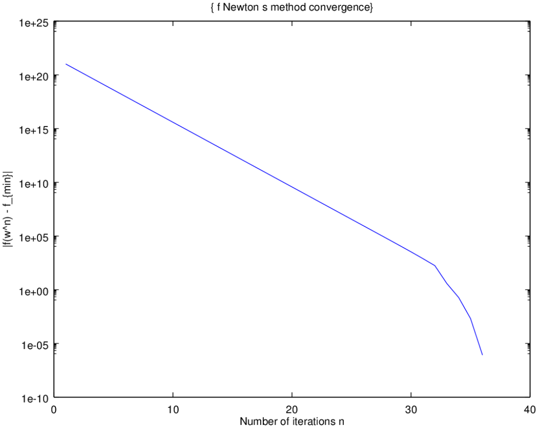
$$\begin{aligned}\Phi(\lambda) &= \sum_i (-log(C - \lambda_i) - log(\lambda_i)) \\ &= \sum_i log(\frac{1}{(C-\lambda_i)\lambda_i}) \\ &= -\sum_i log((C - \lambda_i)\lambda_i)\end{aligned}$$

Problème d'optimisation final

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

Détails de l'implémentation

Tracé de la convergence de la méthode de Newton



Détails de l'implémentation

Dépendance en la taille de l'échantillon

Figure: Tableau récapitulatif

Test	C	d	n	N it.	Temps (s)	Meilleur C	Echec (%)
1	1	40	10	11	25,414	1	0 (*)
1	5	40	10	11	0,177	1	0 (*)
1	10	40	10	11	0,168	1	0 (*)

Détails de l'implémentation

Accélération de la convergence quand C augmente

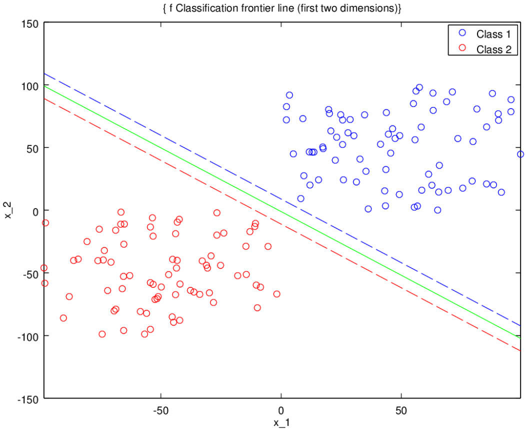
Figure: Tableau récapitulatif

1	5	40000	10	11	0,315
1	5	40	100	12	0,715
1	5	40	1000	?	¿ 10

Tracé de la frontière de classification

Pour $C = 5, n = 150, d = 200$

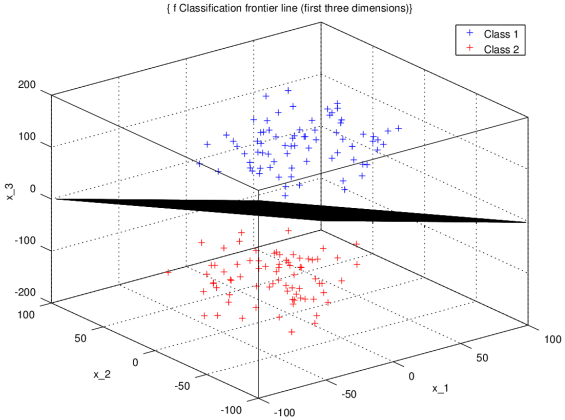
Points centrés réduits avec des fonctions gaussiennes (2D) :



Tracé de la frontière de classification

Pour $C = 5, n = 150, d = 200$

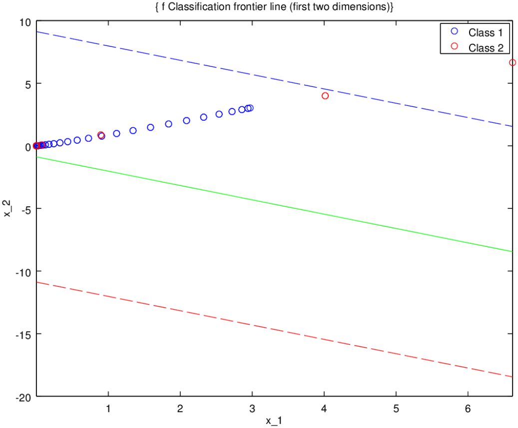
Points centrés réduits avec des fonctions gaussiennes (3D) :



Tracé de la frontière de classification

Pour $C = 5, n = 150, d = 200$

Génération avec des fonctions gaussiennes (2D) :



Tracé de la frontière de classification

Pour $C = 5, n = 150, d = 200$

Génération avec des fonctions gaussiennes (3D) :