

Projet Optimisation : Support Vector Machine

K. Kamtue & Cl. Réda

ENS Cachan

January 1, 2017

1 Description du projet

- Sujet
- Le problème d'optimisation
- Implémentation

2 Résultats

- Détails de l'implémentation
- Tracé de la frontière de classification

1 Description du projet

■ Sujet

- Le problème d'optimisation
- Implémentation

2 Résultats

- Détails de l'implémentation
- Tracé de la frontière de classification

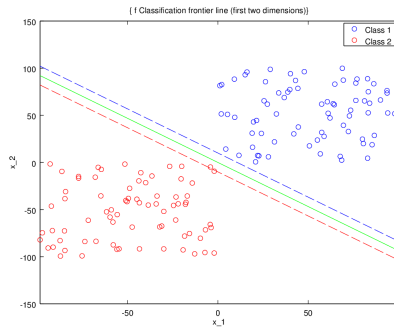
Sujet

Support Machine Vector

Objectif : Faire de l'apprentissage supervisé

- Appliqué à la classification binaire

Figure: Exemple avec deux classes (rouge et bleue)



Le problème d'optimisation

Recherche du problème d'optimisation

Naïvement

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

$$\max_w \gamma = \frac{2}{\|w\|}$$

avec $\forall i, y_i f(x_i) \geq 1$

$$\Leftrightarrow \min_w \|w\|$$

avec $\forall i, y_i f(x_i) \geq 1$

$$\Leftrightarrow \min_w \frac{1}{2} \|w\|^2$$

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 :

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

1 Description du projet

- Sujet
- Le problème d'optimisation
- Implémentation

2 Résultats

- Détails de l'implémentation
- Tracé de la frontière de classification

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

$$\max_{\lambda \in \mathbb{R}^{+m}} -\frac{1}{2} \left\| \sum_i \lambda_i y_i x_i \right\|_2^2 + \mathbf{1}^T \lambda$$

avec $\forall i, 0 \leq \lambda_i \leq C$ 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 :

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

Implémentation

Supprimer les contraintes d'inégalité

Utilisation de la *méthode de la barrière logarithmique* :

Fonction barrière pour éliminer les contraintes d'inégalité

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

Problème d'optimisation final

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

1 Description du projet

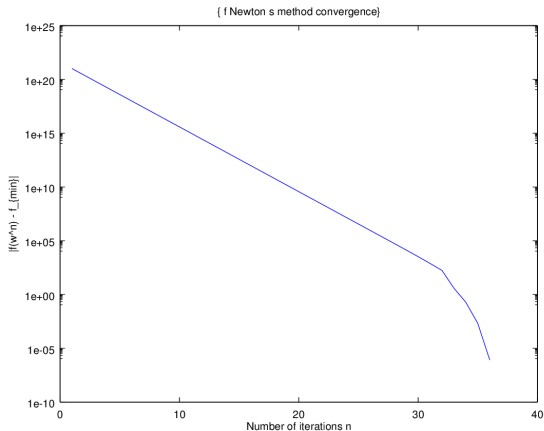
- Sujet
- Le problème d'optimisation
- Implémentation

2 Résultats

- Détails de l'implémentation
- Tracé de la frontière de classification

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

1 Description du projet

- Sujet
- Le problème d'optimisation
- Implémentation

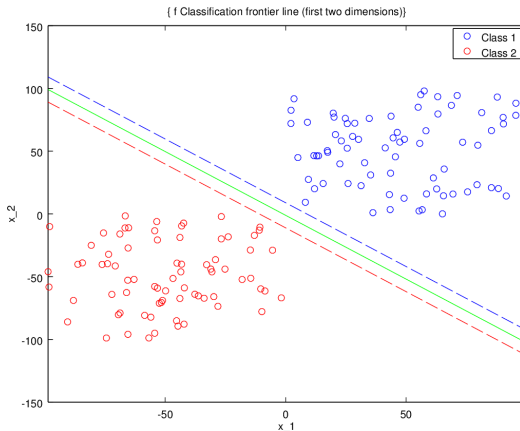
2 Résultats

- Détails de l'implémentation
- Tracé de la frontière de classification

Tracé de la frontière de classification

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

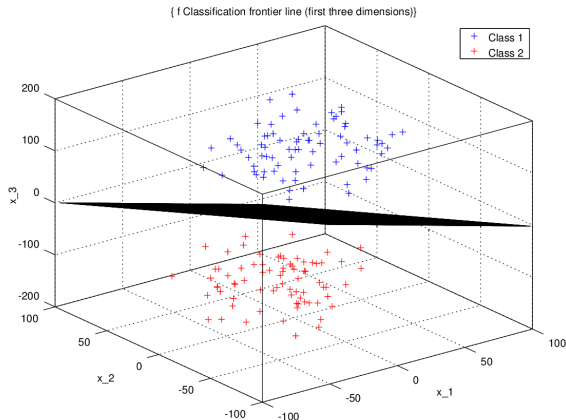
Points centrés réduits avec des fonctions gaussiennes (2D) :



Tracé de la frontière de classification

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

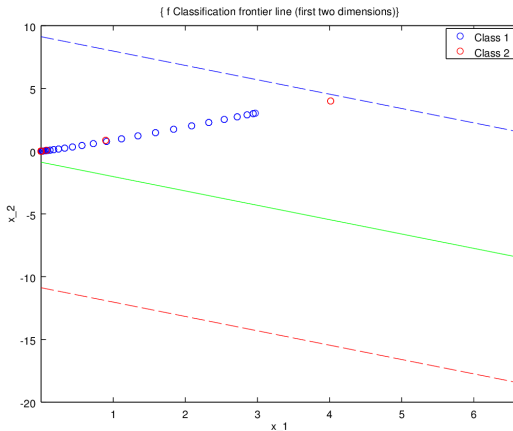
Points centrés réduits avec des fonctions gaussiennes (3D) :



Tracé de la frontière de classification

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

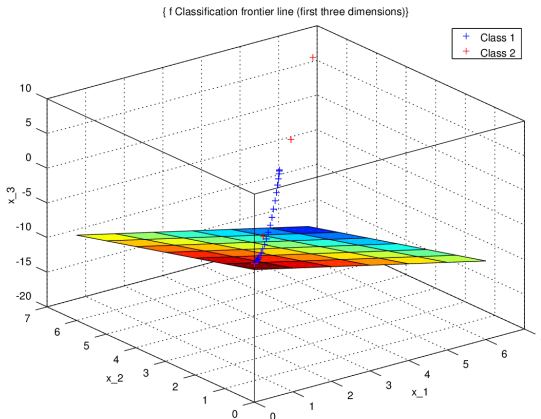
Génération avec des fonctions gaussiennes (2D) :



Tracé de la frontière de classification

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

Génération avec des fonctions gaussiennes (3D) :



Extensions

Ajouts au projet

- Validation croisée (choix de la meilleure valeur de C);
- Implémentation de Coordinate Descent;