

Rapport sur le projet d'Optimisation Support-Vector Machines

Kawisorn Kamtue & Clémence Réda

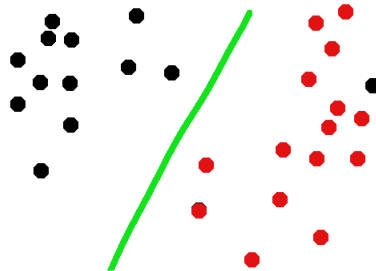
December 4, 2016

1 Support Vector Machine

Les *Support Vector Machine solvers* (SVM) sont une catégorie d'algorithmes d'apprentissage statistique supervisé. Ils permettent de résoudre le problème de classification binaire suivant :

Etant donnés $(x_i)_{i \leq m}$ des points dans \mathbb{R}^n , et $(y_i)_{i \leq m}$ les étiquettes des points tels que l'étiquette de x_i soit $y_i \in \{-1, 1\}$, on cherche la droite qui sépare "le mieux possible" les points dans différentes classes, autrement dit, la frontière de Voronoi entre les deux classes.

Figure 1: Exemple de frontière pour deux classes : celles des points noirs et celle des points rouges



La frontière que l'on recherche est une fonction linéaire, donc de la forme (avec un paramètre ω) :

$$f : X \rightarrow \omega^T X$$

telle que :

$$\begin{aligned} \forall i, y_i = -1 &\Rightarrow f(x_i) \leq -1 \\ \forall i, y_i = 1 &\Rightarrow f(x_i) \geq 1 \\ \Leftrightarrow \forall i, y_i \times f(x_i) &\geq 1 \quad (1) \end{aligned}$$

Pour obtenir un résultat robuste, on souhaite que les deux droites $f(X) = 1$ et $f(X) = -1$ soient les plus distantes possibles. En effet, si ces deux droites sont trop proches, cela signifie que la probabilité d'erreur quant à la prédiction de la classe d'un point proche de ces droites sera importante.

La distance γ entre ces deux droites se calcule de la façon suivante : soient u, v deux points tels que $f(v) = 1$ et $f(u) = -1$. Alors :

$$f(v) - f(u) = \omega \times (v - u) = \omega \times \gamma = 1 - (-1) = 2$$

Finalement, le problème d'optimisation à résoudre pourrait être :

$$\begin{aligned} \max_w \gamma &= \frac{2}{\|\omega\|} \text{ avec (1)} \Leftrightarrow \min_w \|\omega\| \text{ avec (1)} \\ &\Leftrightarrow \min_w \frac{1}{2} \times \|\omega\|^2 \text{ avec (1) pour faciliter les calculs} \end{aligned}$$

Un autre problème se pose si on s'arrête ici : par exemple, dans l'exemple de la frontière de Voronoi que l'on a vu ci-dessus, il n'existe pas de droite telle qu'il n'y ait que de points noirs d'un côté et que des points rouges de l'autre côté, ce qui contredit la condition (1). Pourtant, la droite dessinée en vert peut sembler acceptable comme frontière pour cet ensemble de points.

On tient compte de cette erreur en introduisant les variables $(z_i)_{i \leq m}$. Pour que la condition (1) soit toujours vérifiée, il faut que quand $y_i \times f(x_i) \geq 1$, $z_i = 0$ et lorsque $y_i \times f(x_i) < 1$, $z_i = 1 - y_i \times f(x_i)$. Le but étant de minimiser le nombre de ces erreurs, on utilise un paramètre C constant qui permet d'insister plus ou moins sur la minimisation de ces erreurs :

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

Les fonctions que l'on a introduites sont toutes convexes. Si la dimension des points x_i est petite, nous allons pouvoir utiliser la méthode de Newton pour résoudre ce problème.

2 Calcul du dual

Calculons le lagrangien du problème (P). Soit λ le multiplicateur de Lagrange :

$$\begin{aligned} \forall \omega, \lambda \in \mathbb{R}^{2m}, L(\omega, \lambda, z) &= \\ &= \frac{1}{2} \|\omega\|^2 + C \times \sum_{i \leq m} z_i - \sum_{i < m} \lambda_{m+i} \times z_i \\ &\quad + \sum_{i < m} \lambda_i \times (1 - z_i - y_i \omega^T x_i) \end{aligned}$$

Minimisons L par rapport à ω :

$$\begin{aligned}
\nabla_{\omega} L(\omega, \lambda, z) &= \sum_i w_i + 0 + \sum_{i < m} \lambda_i \times (0 \\
&\quad - 0 - y_i \times x_i) = 0 \\
&\Leftrightarrow \sum_i w_i - \sum_{i < m} \lambda_i \times y_i \times x_i = 0 \\
&\Leftrightarrow \sum_i w_i = \sum_{i < m} \lambda_i \times y_i \times x_i \quad (1)
\end{aligned}$$

Minimisons L par rapport à z :

$$\begin{aligned}
\nabla_z L(\omega, \lambda, z) &= 0 + C \times m - \sum_{i < m} \lambda_{m+i} \\
&\quad + \sum_{i < m} \lambda_i \times (0 - 1 - 0) = 0 \\
&\Leftrightarrow C \times m - \sum_{i < m} \lambda_{m+i} - \sum_{i < m} \lambda_i = 0 \\
&\Leftrightarrow C \times m = \sum_{i < 2m} \lambda_i \quad (2)
\end{aligned}$$

On note alors, en injectant les valeurs de ω et de z ($z \leq 0$ et $C, \lambda_i \geq 0 \forall i$, donc on prend $z = 0$) dans le lagrangien :

$$g(\lambda) = -\frac{1}{2} \times \sum_{i,j} \lambda_i y_i \lambda_j y_j x_i^T x_j$$

Le problème dual est donc :

$$\begin{aligned}
&\min_{\lambda \in \mathbb{R}^m} g(\lambda) \\
&\text{avec } \forall i, 0 \leq \lambda_i \leq C \text{ (vient de (2))}
\end{aligned}$$