Master 1 Informatique – ENSICAEN 2A

Module « Apprentissage »

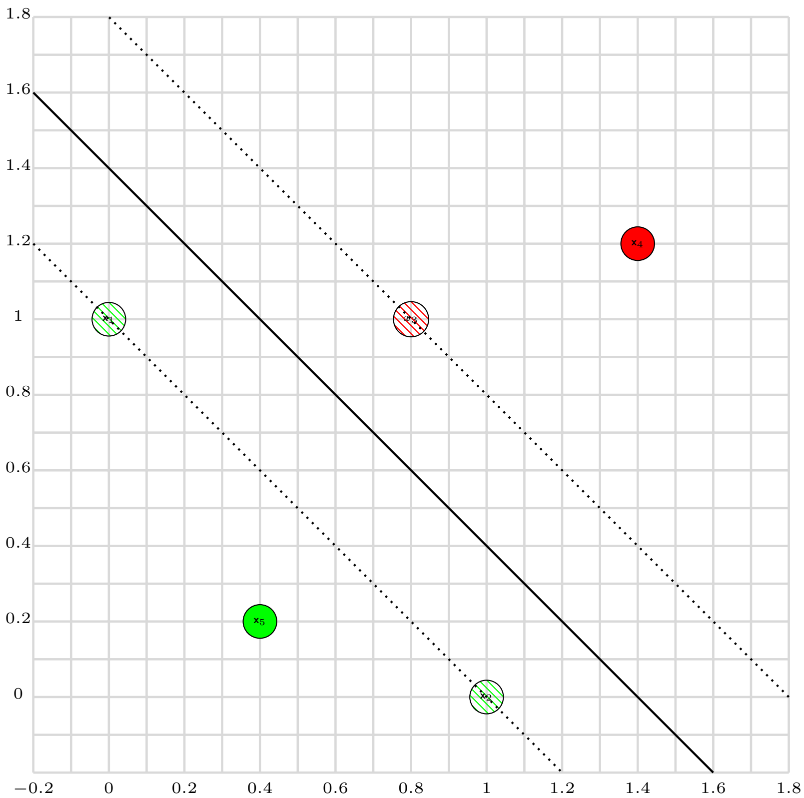
TD numéro 3

Apprentissage supervisé - SVM

1 Support Vector Machine linéaire sans marges floues

On dispose d'un ensemble de points 2D labelisés $\mathcal{X} = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_5, y_5)\}$ définis par $\mathbf{x}_1 = (0, 1)$, $\mathbf{x}_2 = (1, 0)$ $\mathbf{x}_3 = (0.8, 1)$, $\mathbf{x}_4 = (1.4, 1.2)$, $\mathbf{x}_5 = (0.4, 0.2)$ et $y_1 = -1$, $y_2 = -1$, $y_3 = 1$, $y_4 = 1$, $y_5 = -1$. On souhaite classer ces données avec un SVM en utilisant un noyau linéaire $K(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T\mathbf{x}'$ sans marge floue.

1. Sur une figure, disposez les données $\mathcal{X}$, l'hyperplan de séparation des deux classes, les marges géométriques maximales et encerclez les vecteurs de support.
2. Donnez les valeurs des poids $\mathbf{w}$ et du biais b correspondant à l'hyperplan de séparation.
3. Calculez les valeurs des multiplicateurs de Lagrange α_i correspondant à l'hyper-plan de séparation. On rappelle que les $\alpha_i \geq 0$ et $\sum_{i=1}^{N} \alpha_i y_i = 0$.



1. On a trois vecteurs de support : $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$. Cela nous donne trois équations et trois inconnues:

$$\begin{aligned}
f(\mathbf{x}_1) = \quad & w_1x_1^1 + w_2x_1^2 + b = & w_2 + b = -1(EQ1) \\
f(\mathbf{x}_2) = \quad & w_1x_2^1 + w_2x_2^2 + b = & w_1 + b = -1(EQ2) \\
f(\mathbf{x}_1) = \quad & w_1x_3^1 + w_2x_3^2 + b = & 0.8w_1 + w_2 + b = 1(EQ3)
\end{aligned}$$

$EQ3-EQ1-0.8\times EQ2$: $(0.8-0.8)w_1+(1-1)w_2+(1-1-0.8)b = 1+1+0.8 \Rightarrow b = -3.5(EQ4)$
$EQ1 - EQ4$: $w_2 + b - b = -1 + 3.5 \Rightarrow w_2 = 2.5$
$EQ2 - EQ4$: $w_1 + b - b = -1 + 3.5 \Rightarrow w_1 = 2.5$
On a donc $\mathbf{w} = (2.5, 2.5)$ et $b = -3.5$.

2. On a $f(\mathbf{x}) = \sum_{i=1}^{N} y_i\alpha_i\mathbf{x}_i^T\mathbf{x} + b$ avec $\alpha_i \geq 0$ et $\sum_{i=1}^{N} \alpha_i y_i = 0$.
Il n'y a que trois vecteurs de support qui ont leur $\alpha_i \neq 0$. On calcule $f(\mathbf{x})$ pour $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$. Avant il faut calculer les combinaisons de produits scalaires :

$$\begin{aligned}
\mathbf{x}_1^T\mathbf{x}_1 &= & 1 \\
\mathbf{x}_1^T\mathbf{x}_2 &= & 0 \\
\mathbf{x}_1^T\mathbf{x}_3 &= & 0.8 \\
\mathbf{x}_2^T\mathbf{x}_1 &= & 0 \\
\mathbf{x}_2^T\mathbf{x}_2 &= & 1 \\
\mathbf{x}_2^T\mathbf{x}_3 &= & 1 \\
\mathbf{x}_3^T\mathbf{x}_1 &= & 0.8 \\
\mathbf{x}_3^T\mathbf{x}_2 &= & 1 \\
\mathbf{x}_3^T\mathbf{x}_3 &= & 1.64
\end{aligned}$$

$f(\mathbf{x}_1) = \alpha_1(-1)1 + \alpha_2(-1)0 + \alpha_3(1)0.8 - 3.5 = -1$ qui donne $-\alpha_1 + 0.8\alpha_3 = 2.5$
$f(\mathbf{x}_2) = \alpha_1(-1)0 + \alpha_2(-1)1 + \alpha_3(1)1 - 3.5 = -1$ qui donne $-\alpha_2 + \alpha_3 = 2.5$
$f(\mathbf{x}_3) = \alpha_1(-1)0.8 + \alpha_2(-1)1 + \alpha_3(1)1.64 - 3.5 = 1$ qui donne $-0.8\alpha_1 - \alpha_2 + 1.64\alpha_3 = 4.5$
On en déduit que $\alpha_1 = 0.8\alpha_3 - 2.5$ et $\alpha_2 = \alpha_3 - 2.5$.
Comme $\sum_{i=1}^{N} \alpha_i y_i = 0$, on a $\alpha_1(-1) + \alpha_2(-1) + \alpha_3(1) = 0$ c-à-d $-\alpha_1 - \alpha_2 + \alpha_3 = 0$.
On soustrait ceci à $-0.8\alpha_1 - \alpha_2 + 1.64\alpha_3 = 4.5$:
$(-0.8 + 1)\alpha_1 + (-1 + 1)\alpha_2 + (1.64 - 1)\alpha_3 = 4.5$ soit
$0.2\alpha_1 + 0.64\alpha_3 = 4.5$ $0.2(0.8\alpha_3 - 2.5) + 0.64\alpha_3 = 4.5$, soit
$0.8\alpha_3 + 0.5 = 4.5$ et $\alpha_3 = \frac{5}{0.8} = 6.25$.
On en déduit $\alpha_1 = 0.8 \times 6.25 - 2.5 = 2.5$, $\alpha_2 = 6.25 - 2.5 = 3.75$

2 Support Vector Machine linéaire avec marges floues

On dispose d'un ensemble de points 2D labelisés $\mathcal{X}$ et on utilise un SVM à marges floues pour les classifier. Après résolution du problème dual, on obtient les multiplicateurs de Lagrange. Ils sont donnés dans le tableau suivant.

1. Quels sont les vecteurs de support ?
   Ce sont les points de l'ensemble d'apprentissage qui ont leur $\alpha_i \neq 0$ c-à-d $(1.0, 4.0)$, $(4.0, 6.0)$, $(4.5, 3.0)$, $(1.5, 5.0)$, $(2.0, 2.0)$, $(5.0, 3.0)$, $(5.0, 4.5)$.

x_i^1	x_i^2	y_i	α_i
1.0	4.0	+1	2.0
1.5	7.0	+1	0.0
2.0	6.0	+1	0.0
4.0	6.0	+1	1.5
4.5	3.0	+1	2.0
1.5	5.0	-1	2.0
2.0	2.0	-1	1.1
3.0	1.0	-1	0.0
4.0	1.0	-1	0.0
5.0	3.0	-1	0.4
5.0	4.5	-1	2.0

x_i^1	x_i^2	y_i
2.3	3.4	+1
4.4	3.4	-1
5.1	8.1	+1
3.3	3.8	-1
1.7	4.8	-1

Table 1: Gauche : données d'apprentissage labelisées $((x_i^1, x_i^2), y_i)$ et leur multiplicateur de Lagrange α_i. Droite: données de test labelisées $((x_i^1, x_i^2), y_i)$.

2. Le biais est $b = -1.8$. Calculez le vecteur **w**.
   On a $\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i$. On en déduit que $w_1 = 2 \cdot 1 \cdot 1 + 1.5 \cdot 1 \cdot 4 + 2 \cdot 1 \cdot 4.5 + 2 \cdot -1 \cdot 1.5 + 1.1 \cdot -1 \cdot 2 + 0.4 \cdot -1 \cdot 5 + 2 \cdot -1 \cdot 5 = -0.2$
   $w_2 = 2 \cdot 1 \cdot 4 + 1.5 \cdot 1 \cdot 6 + 2 \cdot 1 \cdot 3 + 2 \cdot -1 \cdot 5 + 1.1 \cdot -1 \cdot 2 + 0.4 \cdot -1 \cdot 3 + 2 \cdot -1 \cdot 4.5 = 0.6$
   $\mathbf{w} = (-0.2, 0.6)$

3. À quel point l'hyperplan de séparation coupe l'axe des ordonnées ($x_i^1 = 0$).
   $(-0.2, 0.6)^T \cdot (0, x_i^2) - 1.8 = 0$ et $x_i^2 = 3$.

4. On considère les points $(1, 4)$, $(4.5, 3)$, $(2, 2)$ et $(3, 1)$ (lignes 1, 5, 7 et 8 des données d'apprentissage de la table 1.) Calculez les variables de relâche (slack variables à ξ_i pour ces points et classez ces points parmi (1) bien classé, hors de la marge, (2) bien classé, sur l'hyperplan, (3) bien classé, dans la marge, (4) mal classé.
   On doit avoir $\alpha_i \left(y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 + \xi_i\right) = 0$. On en déduit :

   - $\xi_{(1,4)} = 0.6$ bien classé dans la marge
   - $\xi_{(4.5,3)} = 1.9$ mal classé
   - $\xi_{(2,2)} = 0$ bien classé et sur l'hyperplan
   - $(3, 1)$ n'est pas un vecteur de support, pas de ξ_i à définir.

5. Classez le point $(4, 4)$ en utilisant le SVM.
   $y = \text{sign}(\mathbf{w}^T\mathbf{x} + b) = \text{sign}((-0.2, 0.6)^T \cdot (4, 4) - 1.8) = -1$

6. On dispose de données de test. Quelle est le taux de classification correcte ?

x_i^1	x_i^2	y_i	$f(\mathbf{x}_i)$
2.3	3.4	+1	-1
4.4	3.4	-1	-1
5.1	8.1	+1	+1
3.3	3.8	-1	-1
1.7	4.8	-1	+1

3 points sur 5 sont bien classés le taux est de $\frac{3}{5} = 60\%$

3 Noyaux

On considère le noyau polynomial suivant:

$$K(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T\mathbf{x}' + 1)^2$$

Trouvez l'espace de projection $\Phi : \mathcal{X} \to \mathbb{R}^d$ tel que $K(\mathbf{x}, \mathbf{x}') =< \Phi(\mathbf{x}), \Phi(\mathbf{x}') >$. Quelle est la dimension de cet espace ?
$K(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T\mathbf{x}' + 1)^2 = (x_1x_1' + x_2x_2' + 1)(x_1x_1' + x_2x_2' + 1) =$
$x_1^2x_1'^2 + x_2^2x_2'^2 + 2x_1x_2x_1'x_2 + 2x_1x_1' + 2x_2x_2' + 1$

On en déduit que $\Phi(\mathbf{x}) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ 1 \end{bmatrix} \in \mathbb{R}^6$

4 Support Vector Regression

On dispose d'un ensemble de points labelisés $\mathcal{X} = \{(\mathbf{x}_i, y_i)\}_{i=1,\cdots,N}$ avec $\mathbf{x}_i \in \mathbb{R}^d$ et $y_i \in \mathbb{R}$. On suppose que y_i peut être approché par régression ε près par un SVM $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$: $f(\mathbf{x}_i) - \varepsilon \leq y_i \leq f(\mathbf{x}_i) + \varepsilon$. Ce principe se nomme SVR (Support Vector Regression). On peut trouver $\mathbf{w}$ et b en résolvant $\arg\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2$ sous les contraintes $\mathbf{w}^T\mathbf{x} + b - y_i \leq \varepsilon$ et $-(\mathbf{w}^T\mathbf{x} + b) + y_i \leq \varepsilon$.

1. Donnez la forme duale de ce problème d'optimisation
   $L = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i(\varepsilon - y_i + \mathbf{w}^T\mathbf{x}_i + b) - \sum_{i=1}^{N} \hat{\alpha}_i(\varepsilon + y_i - \mathbf{w}^T\mathbf{x}_i - b)$ avec $\alpha_i \geq 0$ et $\hat{\alpha}_i \geq 0$.
   $\frac{\partial L}{\partial b} = -\sum_{i=1}^{N}(\alpha_i - \hat{\alpha}_i) = 0$
   $\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{N}(\alpha_i - \hat{\alpha}_i)\mathbf{x}_i = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{N}(\alpha_i - \hat{\alpha}_i)\mathbf{x}_i$
   La forme duale est
   $L = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \varepsilon\sum_{i=1}^{N}(\alpha_i + \hat{\alpha}_i) + \sum_{i=1}^{N} y_i(\alpha_i - \hat{\alpha}_i) - \sum_{i=1}^{N}(\alpha_i - \hat{\alpha}_i)(\mathbf{w}^T\mathbf{x}_i + b) =$
   $= \frac{1}{2}\mathbf{w}^T\mathbf{w} - \varepsilon\sum_{i=1}^{N}(\alpha_i + \hat{\alpha}_i) + \sum_{i=1}^{N} y_i(\alpha_i - \hat{\alpha}_i) - \sum_{i=1}^{N}(\alpha_i - \hat{\alpha}_i)(\mathbf{w}^T\mathbf{x}_i) - b\sum_{i=1}^{N}(\alpha_i - \hat{\alpha}_i)$
   $= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j)\mathbf{x}_i^T\mathbf{x}_j - \sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j)\mathbf{x}_i^T\mathbf{x}_j - \varepsilon\sum_{i=1}^{N}(\alpha_i + \hat{\alpha}_i) + \sum_{i=1}^{N} y_i(\alpha_i - \hat{\alpha}_i)$
   $= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j)\mathbf{x}_i^T\mathbf{x}_j - \varepsilon\sum_{i=1}^{N}(\alpha_i + \hat{\alpha}_i) + \sum_{i=1}^{N} y_i(\alpha_i - \hat{\alpha}_i)$

2. Ecrivez f en fonction des variables duales $f(\mathbf{x}) = \sum_{i=1}^{N}(\alpha_i - \hat{\alpha}_i)\mathbf{x}_i^T\mathbf{x} + b$

3. Pour quels points les multiplicateurs de Lagrange sont-ils nuls ? Les conditions KKT imposent que $\alpha_i\left[\varepsilon - y_i + \mathbf{w}^T\mathbf{x}_i + b\right] = 0$ et $\hat{\alpha}_i\left[\varepsilon + y_i - \mathbf{w}^T\mathbf{x}_i - b\right] = 0$. Donc $\alpha_i \neq 0$ ssi $y_i - \mathbf{w}^T\mathbf{x}_i - b = \varepsilon$ et $\hat{\alpha} \neq 0$ ssi $y_i - \mathbf{w}^T\mathbf{x}_i - b = -\varepsilon$. Les points dont y_i est à une distance ε de $f(\mathbf{x}_i)$ sont vecteurs de support (cela définit un tube de rayon 2ε autour de f).

4. Peut-on appliquer l'astuce du noyau au SVR ? Oui car la forme duale n'exploite que des produits scalaires ainsi que $f(x)$.