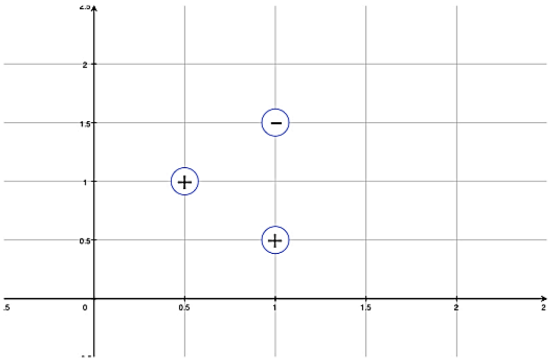
Exercice 1 : (Examen Final 2018)

1. Quel est le but de l'algorithme SVM? Quand peut-il être appliqué avec succès?
2. Si les exemples d'apprentissage sont linéairement séparables, combien de limites de décision peuvent séparer les points de données positifs des points de données négatifs? Quelle limite de décision l'algorithme SVM calcule-t-il? Pourquoi?
3. Nous savons que les frontières de décision ne sont pas linéaires pour la plupart des ensembles de données réelles. Comment traiter cette non-linéarité est-elle par les SVM?
4. Résumer les principaux avantages et limites des algorithmes SVM.
5. Considérons les trois vecteurs d'entrée bidimensionnels linéairement séparables de la figure suivante. Trouvez le SVM linéaire qui sépare de manière optimale les classes en maximisant la marge.



6. Démontrer pour la fonction du noyau polynomial $K(\mathrm{x},\mathrm{z})$ suivante

$$K(\mathrm{x},\mathrm{z}) = (\langle \mathrm{x} \cdot \mathrm{z} \rangle + v)^d, d = 2, v = 1, \mathrm{x} = (x_1, x_2), \mathrm{z} = (z_1, z_2)$$

que :

$$K(\mathrm{x},\mathrm{z}) = \langle \Phi(\mathrm{x}) \cdot \Phi(\mathrm{z}) \rangle \text{, avec } \Phi(\mathrm{y}) = \left(1, \sqrt{2}y_1, \sqrt{2}y_2, {y_1}^2, {y_2}^2, \sqrt{2}y_1 y_2\right)$$

Solution :

1. Les SVM sont des classificateurs linéaires qui cherchent un hyperplan pour séparer deux classes de données, positive et négative. Ils sont applicables avec succès lorsque les deux classes de données de l'ensemble d'apprentissage sont linéairement séparables. Ils conviennent, en particulier pour les données de grande dimension. Les attributs des données d'apprentissage doivent être des nombres réels.
2. Il existe une infinité de limites de décision qui séparent les points de données positifs des points de données négatifs. L'algorithme SVM cherche à trouver la frontière de décision (hyperplan) avec le maximum de marge. Cette frontière minimise la limite supérieure de l'erreur de classification.
3. L'idée est de transformer les données d'entrée non linéairement séparables en un autre espace (généralement de plus grande dimension). Une frontière de décision linéaire peut séparer des

exemples positifs et négatifs dans l'espace transformé. L'espace transformé est appelé l'espace des caractéristiques. L'espace de données d'origine est appelé l'espace d'entrée.

4. Les principaux avantages et limites des algorithmes SVM :

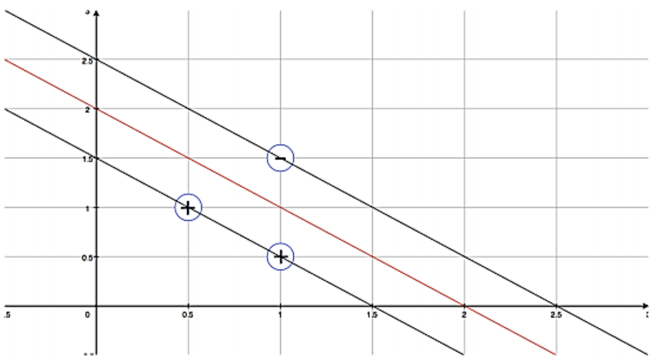
Principaux avantages des algorithmes SVM:

- Ils ont des bases théoriques rigoureuses.
- Effectue la classification plus précisément que la plupart des autres méthodes dans les applications, en particulier pour les données de grande dimension.
- Ils peuvent également être appliqués aux problèmes de classification non linéaire en utilisant les fonctions du noyau (L'astuce des noyaux permet que ces problèmes soient calculables).
- Différents noyaux peuvent être branchés dans le même système d'apprentissage et étudiés indépendamment de celui-ci.

Principales limites des algorithmes SVM:

- Fonctionne uniquement dans un espace réel.
  - Pour un attribut discret, nous devons convertir ses valeurs discrètes en valeurs numériques.
- Ne fait que la classification binaire (à deux classes).
  - Pour les problèmes multi-classes, certaines stratégies peuvent être appliquées, par exemple, la classification un-versus-tous.
- L'hyperplan produit par SVM est difficile à comprendre par les utilisateurs humains. La situation est plus grave dans le cas des noyaux.
  - SVM est couramment utilisé dans des applications qui ne nécessitent pas de compréhension humaine.

5. Les trois points de données sont des vecteurs de support. L'hyperplan de marge $H_{\oplus}$ est la ligne passant par les deux points positifs. L'hyperplan de marge $H_{\ominus}$ est la ligne passant par le point négatif parallèle à $H_{\oplus}$. La frontière de décision est la droite rouge "à mi-chemin" entre $H_{\oplus}$ et $H_{\ominus}$. L'équation de la frontière de décision est -x + 2 = 0. La figure suivante illustre la solution:



6. Démonstration :

$$\boldsymbol{K}(\mathrm{x},\mathrm{z}) = (\langle \mathrm{x}\cdot \mathrm{z}\rangle + 1)^2 = (\langle (x_1,x_2)\cdot(z_1,z_2)\rangle + 1)^2 = (x_1z_1 + x_2z_2 + 1)^2$$
$$= {x_1}^2{z_1}^2 + {x_2}^2{z_2}^2 + 2x_1z_1x_2z_2 + 2x_1z_1 + 2x_2z_2 + 1$$
$$\langle \boldsymbol{\Phi}(\mathrm{x})\cdot\boldsymbol{\Phi}(\mathrm{z})\rangle = \langle \left(1,\sqrt{2}x_1,\sqrt{2}x_2,{x_1}^2,{x_2}^2,\sqrt{2}x_1x_2\right)\cdot\left(1,\sqrt{2}z_1,\sqrt{2}z_2,{z_1}^2,{z_2}^2,\sqrt{2}z_1z_2\right)\rangle$$
$$= 1 + \sqrt{2}x_1\sqrt{2}z_1 + \sqrt{2}x_2\sqrt{2}z_2 + {x_1}^2{z_1}^2 + {x_2}^2{z_2}^2 + \sqrt{2}x_1x_2\sqrt{2}z_1z_2$$
$$= 1 + 2x_1z_1 + 2x_2z_2 + {x_1}^2{z_1}^2 + {x_2}^2{z_2}^2 + 2x_1x_2z_1z_2$$ **CQFD.**

Exercice 2 : (Examen Final 2019)

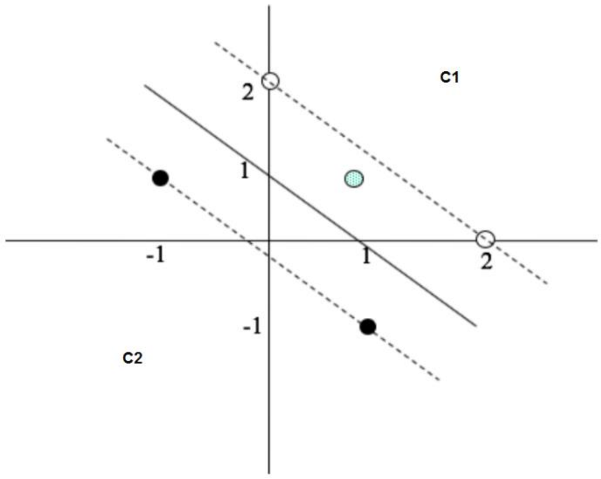
Soient $\boldsymbol{v_1} = (-\mathbf{1}, \mathbf{1}), \boldsymbol{v_2} = (\mathbf{2}, \mathbf{0}), \boldsymbol{v_3} = (\mathbf{1}, -\mathbf{1}), \textbf{and } \boldsymbol{v_3} = (\mathbf{0}, \mathbf{2})$ quatre vecteurs de support définissant une ligne de séparation, et soient les coefficients $\boldsymbol{\alpha}$ et le biais $\boldsymbol{w_0}$:

$$\boldsymbol{\alpha_1 = -0.5, \alpha_2 = 0.5, \alpha_3 = -0.5, \alpha_4 = 0.5 \; et \; w_0 = -1}$$

1. Dessiner un graphe des vecteurs de support et tracez l'hyperplan et la marge qu'ils définissent. (Vous pouvez le faire manuellement sans trouver l'équation de l'hyperplan).
2. Donnez la classification SVM résultante de la nouvelle instance $\mathbf{x} = (\mathbf{1}, \mathbf{1})$.
3. Trouver le vecteur des poids $\boldsymbol{w}$ associé à l'hyperplan de séparation.
4. En utilisant le vecteur des poids $\boldsymbol{w}$ que vous avez obtenus dans la question 3) et le biais $\boldsymbol{w_0}$, trouvez l'équation de l'hyperplan séparateur : $\boldsymbol{w^T}.\mathbf{x} + \boldsymbol{w_0}$. Quelle est la pente de cette ligne ?

Solution :

1. Graphe des vecteurs de support, hyperplan et de la marge :



2. La classification SVM de la nouvelle instance $\mathbf{x} = (\mathbf{1}, \mathbf{1})$:

   1$^{\text{ière}}$ méthode : D'après le graphe ci-dessus, on peut voir directement que la donnée $\mathbf{x}$ appartient à la classe C1.

   2$^{\text{ième}}$ méthode : An appliquant la formule suivante : $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{sgn}\left[\sum_{k=1}^{M} \boldsymbol{\alpha_k y_k (x_k \cdot x) + w_0}\right]$

$$\boldsymbol{f}((\mathbf{1}, \mathbf{1}) = \mathbf{sgn}\,(\mathbf{0.5}((-\mathbf{1}, \mathbf{1}) \cdot (\mathbf{1}, \mathbf{1})) + \mathbf{0.5}((\mathbf{2}, \mathbf{0}) \cdot (\mathbf{1}, \mathbf{1})) + \mathbf{0.5}((\mathbf{1}, -\mathbf{1}) \cdot (\mathbf{1}, \mathbf{1})) + \mathbf{0.5}((\mathbf{0}, \mathbf{2}) \cdot (\mathbf{1}, \mathbf{1})) - \mathbf{1}) = \mathbf{sgn}(\mathbf{0} + \mathbf{1} + \mathbf{0} + \mathbf{1} - \mathbf{1}) = \mathbf{1}$$

3. Le vecteur des poids $\boldsymbol{w}$ associé à l'hyperplan de séparation :

$$\mathbf{0.5}(-\mathbf{1}, \mathbf{1}) + \mathbf{0.5}(\mathbf{2}, \mathbf{0}) + \mathbf{0.5}(\mathbf{1}, -\mathbf{1}) + \mathbf{0.5}(\mathbf{0}, \mathbf{2}) = (\mathbf{1}, \mathbf{1})$$

4. Equation de l'hyperplan séparateur : $\boldsymbol{w^T}.\mathbf{x} + \boldsymbol{w_0}$:
   - $\boldsymbol{x_2 = -x_1 + 1}$
   - **La pente est -1**

Exercice 3 : (Rattrapage 2019)

Soit l'ensemble de données $\mathbf{X} = \{(\mathbf{x}^{(t)}, \mathbf{y}^{(t)}),\ \boldsymbol{t} = \mathbf{1}, .., \mathbf{6}\}$ présenté ci-bas.

$$\mathbf{x}^{(1)} = \begin{bmatrix}7\\5\end{bmatrix}, \mathbf{y}^{(1)} = -1, \qquad \mathbf{x}^{(2)} = \begin{bmatrix}10\\8\end{bmatrix}, \mathbf{y}^{(2)} = -1, \qquad \mathbf{x}^{(3)} = \begin{bmatrix}13\\9\end{bmatrix}, \mathbf{y}^{(3)} = -1,$$

$$\mathbf{x}^{(4)} = \begin{bmatrix}8\\13\end{bmatrix}, \mathbf{y}^{(4)} = 1, \qquad \mathbf{x}^{(5)} = \begin{bmatrix}6\\9\end{bmatrix}, \mathbf{y}^{(5)} = 1, \qquad \mathbf{x}^{(6)} = \begin{bmatrix}2\\10\end{bmatrix}, \mathbf{y}^{(6)} = 1$$

1) Tracer ces points en deux dimensions.

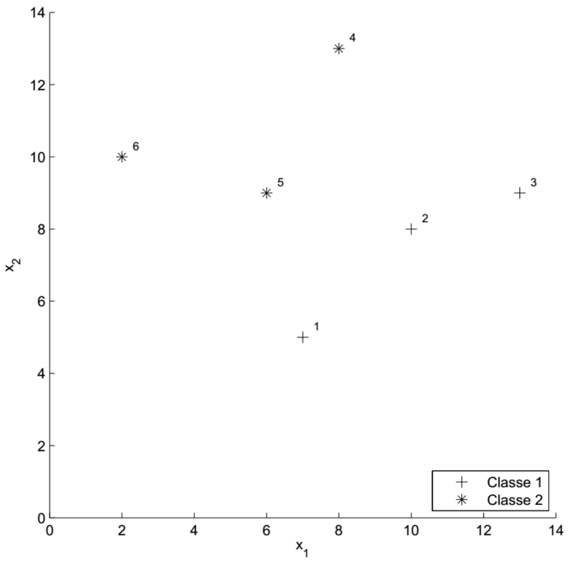
2) Supposons que l'on veut classer ces données avec un classifieur de type SVM utilisant un noyau linéaire $(\boldsymbol{K}(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle)$**.** Tracez les données de l'ensemble $\mathbf{X}$, les marges géométriques maximales obtenues avec le SVM, l'hyperplan séparateur correspondant, et encerclez les données agissant comme vecteurs de support.

3) Donnez les valeurs des poids $\boldsymbol{w}$ et biais $\boldsymbol{w_0}$ correspondant au discriminant linéaire maximisant les marges géométriques tracées en question **2)**.

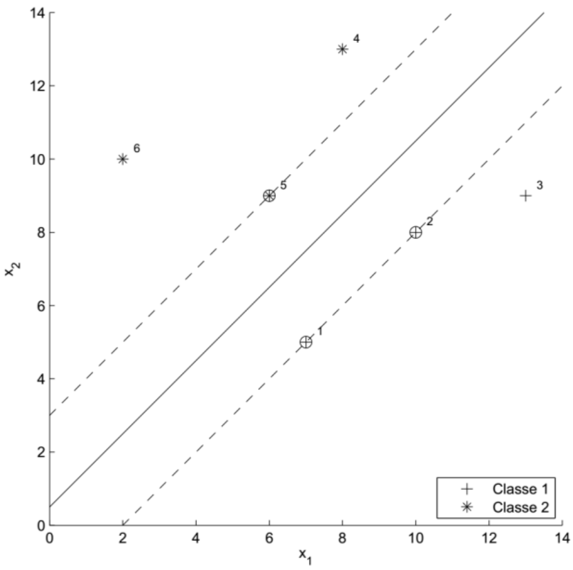
*** Indice :** il n'est pas nécessaire de calculer les $\boldsymbol{\alpha_i}$ pour répondre à la question.

Solution :

1) Tracer ces points en deux dimensions.



2) Supposons que l'on veut classer ces données avec un classifieur de type SVM utilisant un noyau



3) Trois données sont identifiées comme vecteurs de support : x_1, x_2 et x_5. En travaillant directement dans l'espace d'entrée, ceci nous donne trois équations et trois inconnus.

$$h(x^1) = w_2 x_2^1 + w_1 x_1^1 + w_0 = 5w_2 + 7w_1 + w_0 = -1$$

$$h(x^2) = w_2 x_2^2 + w_1 x_1^2 + w_0 = 8w_2 + 10w_1 + w_0 = -1$$

$$h(x^5) = w_2 x_2^5 + w_1 x_1^5 + w_0 = 9w_2 + 6w_1 + w_0 = 1$$

La résolution de ce système d'équation peut se faire par la suivante.

$$(8 \times \mathrm{EQ1}) - (5 \times \mathrm{EQ2}) : (40 - 40)w_2 + (56 - 50)w_1 + (8 - 5)w_0 = -8 + 5$$

$$6w_1 + 3w_0 = -3$$

$$(9 \times \mathrm{EQ2}) - (8 \times \mathrm{EQ3}) : (72 - 72)w_2 + (90 - 48)w_1 + (9 - 8)w_0 = -9 - 8$$

$$42w_1 + w_0 = -17$$

$$\mathrm{L1} - (3 \times \mathrm{L2}) : (6 - 126)w_1 + (3 - 3)w_0 = -3 + 51$$

$$-120w_1 = 48 \Rightarrow w_1 = -0.4$$

$$L1 - (6 \times L3) : (6 - 6)w_1 + (3 - 0)w_0 = -3 - 6(-0.4)$$

$$3w_0 = -3 + 2.4 \Rightarrow w_0 = -0.2$$

$$\mathrm{EQ1} - (7 \times \mathrm{L3}) - \mathrm{L4} : 5w_2 + (7 - 7)w_1 + (1 - 1)w_0 = -1 - 7(-0.4) - (-0.2)$$

$$5w_2 = -1 + 2.8 + 0.2 = 2 \Rightarrow w_2 = 0.4$$

La résolution de ce système d'équations nous donne les valeurs suivantes :

$$w_2 = 0.4, w_1 = -0.4, w_0 = -0.2$$