Centre Universitaire de Mila

Institut: ST, Département: MI

Master 1 – STIC

Corrigé de l'examen final en « Apprentissage Automatique »

Durée de l'examen : 1h30 / Mila le : 16-05-2018

Documents non autorisés

Exercice 1 : Classificateur de Bayes Naïve (10 points)

Supposons que nous ayons l'ensemble de données suivant qui enregistre dans une période de 25 jours si une personne a joué au tennis ou non en fonction des conditions du ciel et de vent.

Date	Ciel	Vent	Jouer au Tennis
1	Ensoleillé	Faible	Non
2	Ensoleillé	Fort	Non
3	Couvert	Faible	Oui
4	Pluie	Faible	Oui
5	Couvert	Faible	Oui
6	Pluie	Fort	Non
7	Couvert	Fort	Oui
8	Ensoleillé	Faible	Non
9	Ensoleillé	Faible	Oui
10	Couvert	Faible	Oui
11	Ensoleillé	Fort	Oui
12	Couvert	Fort	Oui
13	Couvert	Faible	Oui
14	Pluie	Fort	Non
15	Ensoleillé	Fort	Oui
16	Couvert	Fort	Non
17	Couvert	Faible	Oui
18	Pluie	Faible	Non
19	Ensoleillé	Faible	Non
20	Pluie	Fort	Oui
21	Ensoleillé	Faible	Oui
22	Couvert	Faible	Non
23	Couvert	Faible	Oui
24	Ensoleillé	Fort	Oui
25	Couvert	Faible	Non

Nous voulons prédire si la personne jouera au tennis dans les trois jours à venir.

- **Jour 26: (Ciel = Ensoleillé, Vent = Fort) → Jouer au Tennis =?**
- **Jour 27: (Ciel = Couvert, Vent = faible) → Jouer au Tennis =?**
- **Jour 28: (Ciel = Pluie, Vent = Faible) → Jouer au Tennis =?**

Calculez manuellement les prédictions (c'est-à-dire que la personne jouera au tennis ou non) pour les trois jours à venir (c'est-à-dire les jours 26 à 28) en utilisant l'approche de classification de Bayes Naïve (CBN).

Solution

- **Solution pour la date 26 :**

$$HBN = argmax_{h\in\{Oui,Non\}} P(h) \times \prod_{j=1}^{2} P(z_j|h)$$

$$= argmax_{h\in\{Oui,Non\}} P(h) \times P(Ciel = Ensoleillé|h) \times P(Vent = Fort|h)$$

$$P(Oui) = \frac{15}{25} = \frac{3}{5} \qquad P(Non) = \frac{10}{25} = \frac{2}{5}$$

$$P(Ciel = Ensoleillé|Oui) = \frac{5}{15} \qquad P(Vent = Fort|Oui) = \frac{6}{15}$$

$$P(Ciel = Ensoleillé|Non) = \frac{4}{10} \qquad P(Vent = Fort|Non) = \frac{4}{10}$$

$$P(Oui) \times P(Ciel = Ensoleillé|Oui) \times P(Vent = Fort|Oui) = \frac{3}{5} \times \frac{5}{15} \times \frac{6}{15} = 0.08$$

$$P(Non) \times P(Ciel = Ensoleillé|Non) \times P(Vent = Fort|Non) = \frac{2}{5} \times \frac{4}{10} \times \frac{4}{10} = 0.064$$

$HBN(h = Oui) > HBN(h = Non)$, donc Jour 26 → Jouer au Tennis.

- **Solution pour la date 27 :**

$$HBN = argmax_{h\in\{Oui,Non\}} P(h) \times \prod_{j=1}^{2} P(z_j|h)$$

$$= argmax_{h\in\{Oui,Non\}} P(h) \times P(Ciel = Couvert|h) \times P(Vent = Faible|h)$$

$$P(Oui) = \frac{3}{5} \qquad P(Non) = \frac{2}{5}$$

$$P(Ciel = Couvert|Oui) = \frac{8}{15} \qquad P(Vent = Faible|Oui) = \frac{9}{15}$$

$$P(Ciel = Couvert|Non) = \frac{3}{10} \qquad P(Vent = Faible|Non) = \frac{6}{10}$$

$$P(Oui) \times P(Ciel = Couvert|Oui) \times P(Vent = Faible|Oui) = \frac{3}{5} \times \frac{8}{15} \times \frac{9}{15} = 0.192$$

$$P(Non) \times P(Ciel = Couvert|Non) \times P(Vent = Faible|Non) = \frac{2}{5} \times \frac{3}{10} \times \frac{6}{10} = 0.072$$

$HBN(h = Oui) > HBN(h = Non)$, donc Jour 27 → Jouer au Tennis.

- **Solution pour la date 28 :**

$$HBN = argmax_{h\in\{Oui,Non\}} P(h) \times \prod_{j=1}^{2} P(z_j|h)$$

$$= argmax_{h\in\{Oui,Non\}} P(h) \times P(Ciel = Pluie|h) \times P(Vent = faible|h)$$

$$P(Oui) = \frac{3}{5} \qquad P(Non) = \frac{2}{5}$$

$$P(Ciel = Pluie|Oui) = \frac{2}{15} \qquad P(Vent = Faible|Oui) = \frac{9}{15}$$

$$P(Ciel = Pluie|Non) = \frac{3}{10} \qquad P(Vent = Faible|Non) = \frac{6}{10}$$

$$P(Oui) \times P(Ciel = Pluie|Oui) \times P(Vent = Faible|Oui) = \frac{3}{5} \times \frac{2}{15} \times \frac{9}{15} = 0.048$$

$$P(Non) \times P(Ciel = Pluie|Non) \times P(Vent = Faible|Non) = \frac{2}{5} \times \frac{3}{10} \times \frac{6}{10} = 0.072$$

$HBN(h = Non) > HBN(h = Oui)$, donc Jour 28 → Ne pas jouer au Tennis.

Exercice 2 : Machines à vecteurs de support (10 points)

1. Quel est le but de l'algorithme SVM? Quand peut-il être appliqué avec succès?
   Solution
   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. 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?
   Solution
   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. 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?
   Solution
   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. Résumer les principaux avantages et limites des algorithmes SVM.

Solution

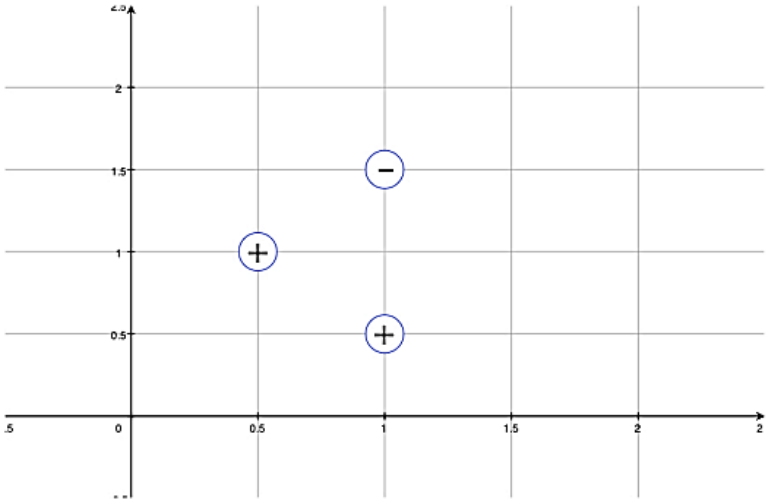
<u>Principaux avantages des algorithmes SVM:</u>

- **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.**

<u>Principales limites des algorithmes SVM:</u>

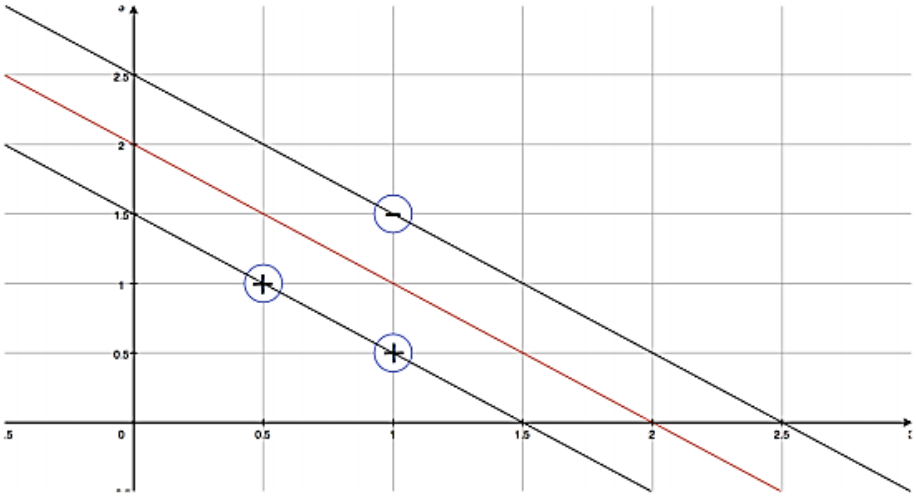
- **Fonctionne uniquement dans un espace réel.**
  - **Pour un attribut discret, nous devons convertir ses valeurs discrètes en valeurs numériques.**
- **Ne fait que la classification binaire (à deux classes).**
  - **Pour les problèmes multi-classes, certaines stratégies peuvent être appliquées, par 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. 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.



Solution

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. Démontrer pour la fonction du noyau polynomial

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

que

$$\boldsymbol{K(\mathbf{x},\mathbf{z}) = \langle \Phi(\mathbf{x})\cdot\Phi(\mathbf{z})\rangle}\textbf{ , avec }\boldsymbol{\Phi(\mathbf{y}) = \left(1, \sqrt{2}y_1, \sqrt{2}y_2, {y_1}^2, {y_2}^2, \sqrt{2}y_1y_2\right)}$$

Solution

$$\boldsymbol{K(\mathbf{x},\mathbf{z}) = (\langle \mathbf{x}\cdot\mathbf{z}\rangle + 1)^2 = (\langle (x_1, x_2)\cdot(z_1, z_2)\rangle + 1)^2 = (x_1z_1 + x_2z_2 + 1)^2}$$
$$\boldsymbol{= {x_1}^2{z_1}^2 + {x_2}^2{z_2}^2 + 2x_1z_1x_2z_2 + 2x_1z_1 + 2x_2z_2 + 1}$$

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

CQFD.