Numéro d'anonymat :

Contrôle de Connaissance
L2 Informatique - Introduction à l'apprentissage automatique- Université Paris Sud

Mardi 7 Janvier 2020
Durée : 2h00
Documents autorisés : supports et notes de cours

Partie I : Apprentissage supervisé

1 Perceptron [11 pts]

On cherche à écrire un algorithme permettant de distinguer les courriels importants des autres. On considérera donc une tâche de classification binaire. Pour entraîner notre algorithme, nous disposons de N données d'apprentissage : $(\boldsymbol{x}_i, \tilde{y}_i)_{i=1,..,N}$ où $\boldsymbol{x}_i$ représente la i-ème données et $\tilde{y}_i$ son étiquette ($+1$ si le courriel est important, -1 sinon).

1.1 Questions de cours [2 pts]

1. Quelle est la fonction d'activation utilisée dans le perceptron binaire qui permet de décider de la classe d'une donnée ?
2. Illuster en deux dimensions, sur un jeu de données avec deux classes, un cas où le perceptron binaire peut réussir à classifier les données et un cas où cela n'est pas possible.
3. Rappeler comment s'expriment les sorties du perceptron multiclasse et expliciter les différents paramètres/variables.
4. Expliquer à l'aide d'un graphe comment on choisit l'époque à laquelle il faut arrêter l'apprentissage.

1.2 Le perceptron binaire "classique" [4 pts]

On s'attaque ici au problème sous l'angle du perceptron binaire non probabiliste. Nous supposerons que nous disposons des données suivantes (au format (x_{i1}, x_{i2}) pour la donnée i) :

$$\begin{cases} \boldsymbol{x}_1 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \\ \tilde{y}_1 = 1 \end{cases}, \begin{cases} \boldsymbol{x}_2 = \begin{pmatrix} 3 \\ -0.5 \end{pmatrix} \\ \tilde{y}_2 = 1 \end{cases}, \begin{cases} \boldsymbol{x}_3 = \begin{pmatrix} -1 \\ -2 \end{pmatrix} \\ \tilde{y}_3 = -1 \end{cases}, \begin{cases} \boldsymbol{x}_4 = \begin{pmatrix} -3 \\ -2 \end{pmatrix} \\ \tilde{y}_4 = -1 \end{cases}$$

et que le vecteur de poids initial est donné par $\mathbf{w'}^t = (w_0, w_1, w_2) = (2, -1, 1)$

1. Indiquer quelles sont les données bien classées et les données mal classées <u>et</u> les représenter sur un graphe.
2. Tracer la fontière de décision correspondant au $\mathbf{w'}^t$ initial.
3. Effectuer une époque sur ces données (en les prenant dans l'ordre que vous voulez) et indiquer la nouvelle valeur des paramètres $\mathbf{w}$ du percetron après avoir vu chaque point.
4. Sur le graphe précédent, représenter la nouvelle frontière de décision.
5. L'algorithme peut-il converger sur ces données ?

1.3 Perceptron multicouche avec la fonction Sign [5 pts]

Sur le précédent problème, on a rencontré en plus des 3 données précédentes aux nouvelles données suivantes :

$$\left\{ \begin{array}{l} \boldsymbol{x}_5 = \begin{pmatrix} 1 \\ -3 \end{pmatrix} \\ \\ \tilde{y}_5 = 1 \end{array} \right. , \left\{ \begin{array}{l} \boldsymbol{x}_6 = \begin{pmatrix} 1 \\ -4 \end{pmatrix} \\ \\ \tilde{y}_6 = 1 \end{array} \right. , \left\{ \begin{array}{l} \boldsymbol{x}_7 = \begin{pmatrix} -8 \\ 1 \end{pmatrix} \\ \\ \tilde{y}_7 = 1 \end{array} \right.$$

On se propose d'utiliser un perceptron multicouche afin de classifier l'ensemble des données. On va étudier le fonctionnement d'un perceptron multicouche avec une couche cachée et on utilisera des fonctions d'activation de type $\mathrm{sign}(x)$ pour la couche intermédiaire. La fonction d'activation sign se comporte de la façon suivante :

$$\mathrm{sign}(x) = \left\{ \begin{array}{ll} 1 & \text{si } x > 0 \\ -1 & \text{si } x < 0 \end{array} \right. \tag{1}$$

1. Expliquer pourquoi il est impossible de classifier ces données à l'aide d'un perceptron classique.
2. Sur un graphique, représenter l'allure de cette fonction d'activation.

On souhaite vérifier qu'il existe une solution à notre problème de classification. Pour cela, on va considérer deux perceptrons simples qui composeront la couche cachée, dont les fonctions d'activation sont données par la fonction sign. On va regarder les deux perceptron séparément avant de conclure. La sortie de chacun des perceptron est obtenue par $y = \mathrm{sign}(w_0 + w_1 x_1 + w_2 x_2)$.

3. On veut trouver les poids d'un perceptron (1) permettant de classer les élements $x_2 > 0$ en $y^{(1)} = +1$ et $x_2 < 0$ en $y^{(1)} = -1$. Donner les poids $w_0^{(1)}, w_1^{(1)}, w_2^{(1)}$ d'un tel perceptron.
4. On veut trouver les poids d'un perceptron (2) permettant de classer les élements $x_1 > 0$ en $y^{(2)} = +1$ et $x_1 < 0$ en $y^{(2)} = -1$. Donner les poids $w_0^{(2)}, w_1^{(2)}, w_2^{(2)}$ d'un tel perceptron.
5. Identifier les valeurs de $y^{(1)}$ et $y^{(2)}$ pour les quatre cas suivants
   (a) $x_1 < 0$ et $x_1 < 0$
   (b) $x_1 > 0$ et $x_1 < 0$
   (c) $x_1 < 0$ et $x_1 > 0$
   (d) $x_1 > 0$ et $x_1 > 0$
6. Donner pour toutes les données, la valeur de $y^{(1)}$ et $y^{(2)}$.
7. Est-il possible de classifier les données dans le plan $y^{(1)}$, $y^{(2)}$? (justifier votre réponse).

Partie II : Apprentissage non-supervisé

2 Partitionnement de données [9 pts]

2.1 Questions de cours [2 pts]

1. Rappeler comment fonctionne l'algorithme des k-moyennes.
2. Quelles sont les paramètres supplémentaires que l'on a lorsque l'on passe à la mixture de gaussiennes ?
3. Comment est fixé le nombre de partitions dans l'algorithme de partitionnement par densité (DBSCAN) ?

2.2 Application [5 pts]

1. Appliquer deux itérations de l'algorithme des k-moyennes (avec calcul si nécessaire) sur le jeu de données de la figure 1 en partant des trois centres suivants : $c_1 = (1, 1)$, $c_2 = (3, 4)$ et $c_3 = (6, 1)$.
2. Que donnerait un partitionnement du même jeu de données en utilisant les deux centres suivants : $c_1 = (2, 3)$ et $c_2 = (6, 2)$ (faire un schéma pour répondre à la question, ne pas faire de calcul).
3. En supposant un partionnement parfait selon 3 centres avec la mixture de gaussiennes, donner une estimation approchée de la probabilité à priori d'appartenir à chacune de partitions (le paramètre ρ des gaussiennes).
4. Expliquer comment il est possible d'utiliser un algorithme de partitionnement pour segmenter une image par couleur.
5. Sachant que les images couleurs sont codées sur 24 bits, quelle facteur de compression (rapport entre la taille non-compressée et la taille compressée de l'image) peut-on obtenir si on utilise 32 centres (donc 32 couleurs) pour segmenter une image par couleur.

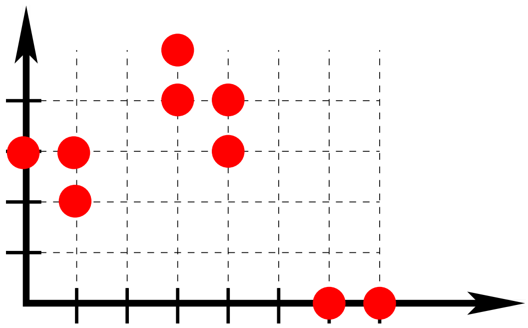

FIGURE 1 – blabla

2.3 Classification à l'aide de la mixture de Gaussiennes [2 pts]

On suppose que l'on a accès à un jeu de données $\{\boldsymbol{x}_i, \tilde{y}_i\}_{i=1,..,N}$ où $\boldsymbol{x}_i$ représente la donnée numéro i et $\tilde{y}_i$ son étiquette, un entier compris en 1 et K.

— Comment peut-on fixer les paramètres des gaussiennes directement ?

— Comment peut-on utiliser la mixture de Gaussiennes ainsi établie pour prédire la classe de nouvelles données ?