Introduction à l'apprentissage automatique (GIF-4101)
Département de génie électrique et de génie informatique
Automne 2022

EXAMEN

Instructions : – Identifiez-vous bien sur la page titre ;
– Répondez directement dans le questionnaire fourni ;
– Une feuille aide-mémoire recto verso manuscrite est permise ;
– Durée de l'examen : 1 h 50.

Pondération : Cet examen compte pour 20% de la note finale.

Prénom : ______________________________

Nom : ______________________________

NI : ______________________________

Signature : ______________________________

GIF-4101

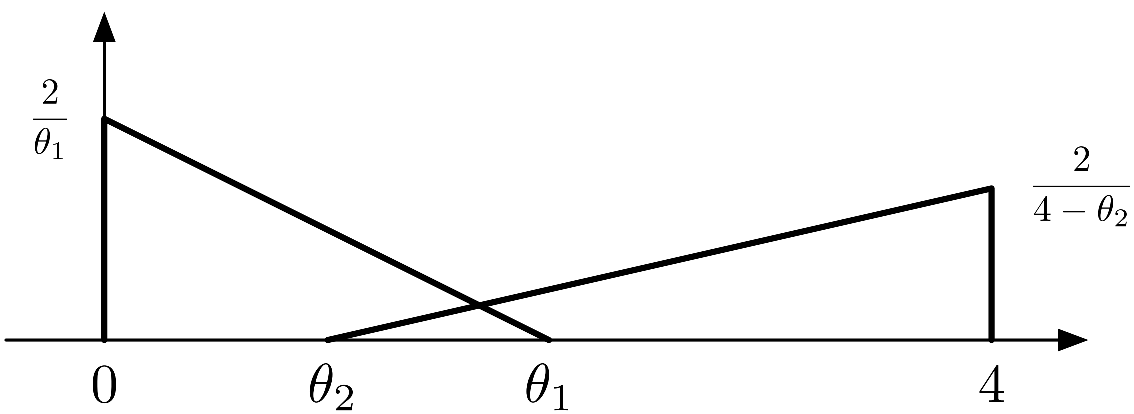
Question 1 (30 points sur 100)

Soit un système de classement paramétrique à deux classes et comportant une variable en entrée. La modélisation des distributions pour chaque classe est donnée par les équations suivantes :

$$p(x|C_1) = \begin{cases} \frac{-2\,(x-\theta_1)}{(\theta_1)^2} & \text{si } x \in [0, \theta_1] \\ 0 & \text{autrement} \end{cases},$$

$$p(x|C_2) = \begin{cases} \frac{2\,(x-\theta_2)}{(4-\theta_2)^2} & \text{si } x \in [\theta_2, 4] \\ 0 & \text{autrement} \end{cases}.$$

Ainsi, la paramétrisation de la distribution de la classe C_1 est donnée par θ_1, alors que celle de la classe C_2 est donnée par θ_2. On fait également l'hypothèse que $0 \leq \theta_2 \leq \theta_1 \leq 4$. La figure suivante présente le tracé de ces distributions de classes.



(10 pts) (a) Supposons que $\theta_1 = 3$ et $\theta_2 = 2$, donnez la fonction $\mathrm{h}(x)$ correspondant à la prise de décision pour le classement de données selon la valeur de $x \in [0, 4]$. Supposez que les probabilités *a priori* des classes sont égales, soit $P(C_1) = P(C_2) = 0{,}5$. Supposez également une perte égale pour les différents types d'erreurs. Donnez les développements menant à votre fonction de décision.

(10 pts) (b) Calculez le taux d'erreur bayésien optimal que l'on obtient avec le classifieur calculé au point précédent. Le taux d'erreur bayésien optimal correspond au taux d'erreur obtenu lorsque les données classées suivent parfaitement les distributions estimées pour le classement.

(10 pts) (c) Supposons maintenant que la fonction de perte est variable selon le type d'erreur que fait notre classifieur. Plus précisément, si une donnée est classée comme étant dans la classe C_2 mais appartient en fait à la classe C_1, la perte est de $\mathcal{L}(\alpha_2,C_1) = 1$, alors que la perte pour une donnée classée comme étant de la classe C_1, mais appartenant en fait à la classe C_2 est de $\mathcal{L}(\alpha_1,C_2) = 0{,}5$. Calculez la nouvelle fonction $\mathrm{h}(x)$ correspondant à la prise de décision pour le classement de données selon cette fonction de perte dans le domaine $x \in [0, 4]$. Supposez que les autres paramètres sont les mêmes qu'aux points précédents, soit que $\theta_1 = 3$, $\theta_2 = 2$ et $P(C_1) = P(C_2) = 0{,}5$. Donnez les développements menant à votre fonction de décision.

Question 2 (40 points sur 100)

Soit un réseau de neurones de type RBF pour deux classes, composé d'une couche cachée de R neurones de type gaussien, suivi d'une couche de sortie d'un neurone avec fonction de transfert linéaire. La valeur de la sortie pour un tel réseau de neurones pour une valeur d'entrée $\mathbf{x}$ est donnée par l'équation suivante,

$$\mathrm{h}(\mathbf{x}) = \sum_{i=1}^{R} w_i \phi_i(\mathbf{x}) + w_0 = \sum_{i=1}^{R} w_i \exp\left[-\frac{\|\mathbf{x}-\mathbf{m}_i\|^2}{2s_i^2}\right] + w_0,$$

où :
— $\mathbf{m}_i$ est la valeur du centre du i-ème neurone gaussien de la couche cachée ;
— s_i est l'étalement du i-ème neurone gaussien ;
— w_i est le poids connectant le i-ème neurone gaussien de la couche cachée au neurone de sortie ;
— w_0 est le poids-biais du neurone de sortie.

Supposons que l'on fixe les étalements s_i à des valeurs prédéterminées et que l'on veut apprendre les valeurs w_i, w_0 et $\mathbf{m}_i$ par descente du gradient, en utilisant comme critère l'erreur quadratique moyenne,

$$E = \frac{1}{2N} \sum_{\mathbf{x}^t \in \mathcal{X}} (e^t)^2 = \frac{1}{2N} \sum_{\mathbf{x}^t \in \mathcal{X}} [r^t - \mathrm{h}(\mathbf{x}^t)]^2,$$

où :
— r^t est la valeur désirée pour le neurone de sortie du réseau ;
— $\mathcal{X}$ est l'ensemble des N données d'entraînement.

(20 pts) (a) Développez les équations permettant de mettre à jour les poids w_i et w_0 du neurone de sortie par descente du gradient, en utilisant le critère de l'erreur quadratique moyenne.

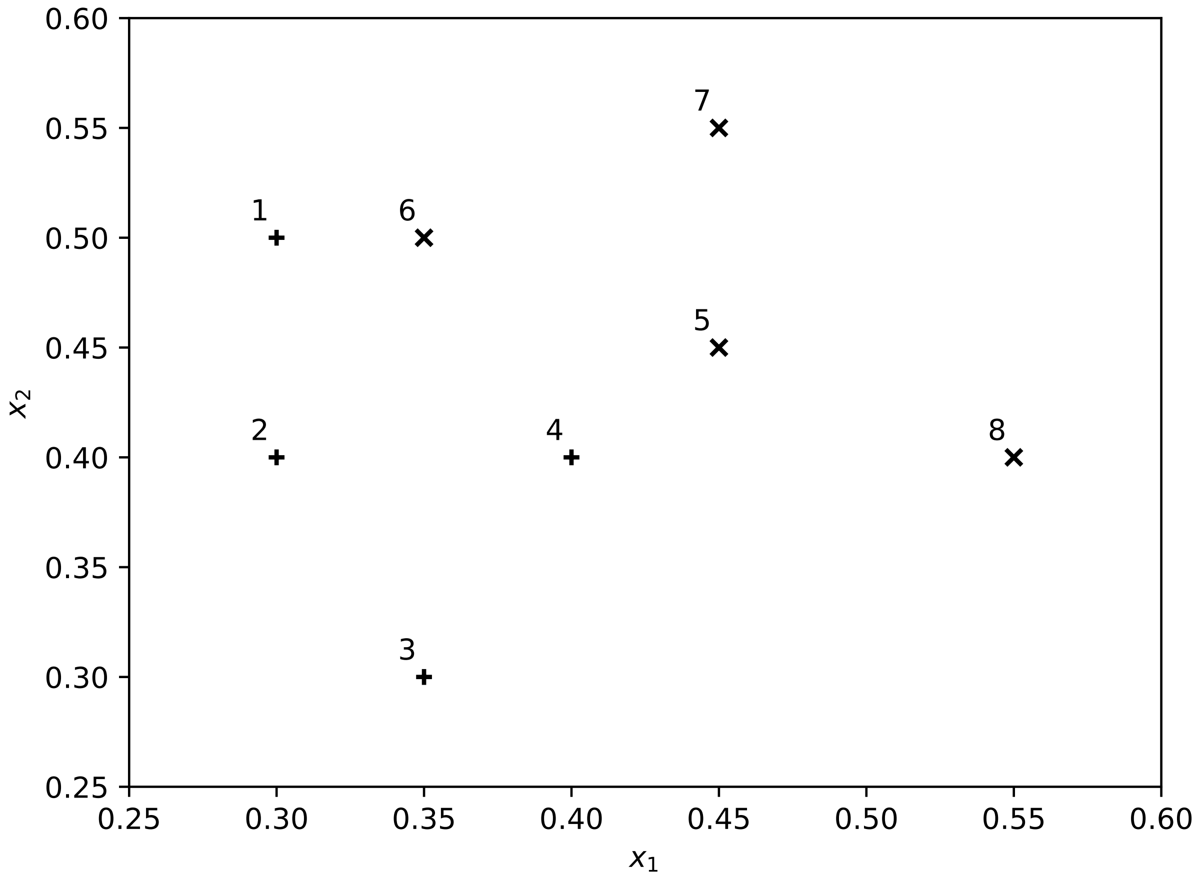
(20 pts) (b) Développez les équations permettant de mettre à jour les valeurs des centres $\mathbf{m}_i$ des neurones gaussiens de la couche cachée par descente du gradient, en utilisant le critère de l'erreur quadratique moyenne.

Question 3 (30 points sur 100)

Soit le jeu de données suivant, en deux dimensions :

$$\mathbf{x}^1 = [0{,}3\ 0{,}5]^\top, \quad \mathbf{x}^2 = [0{,}3\ 0{,}4]^\top, \quad \mathbf{x}^3 = [0{,}35\ 0{,}3]^\top, \quad \mathbf{x}^4 = [0{,}4\ 0{,}4]^\top,$$
$$\mathbf{x}^5 = [0{,}45\ 0{,}45]^\top, \quad \mathbf{x}^6 = [0{,}35\ 0{,}5]^\top, \quad \mathbf{x}^7 = [0{,}45\ 0{,}55]^\top, \quad \mathbf{x}^8 = [0{,}55\ 0{,}4]^\top.$$

Les étiquettes de ces données sont $r^1 = r^2 = r^3 = r^4 = -1$ et $r^5 = r^6 = r^7 = r^8 = 1$.
Le graphique ici bas présente le tracé de ces données.



Nous obtenons le résultat suivant en effectuant l'entraînement d'un SVM linéaire à **marge douce** avec ces données, en utilisant comme valeur de paramètre de régularisation $C = 200$:

$$\alpha^1 = 180, \quad \alpha^2 = 0, \quad \alpha^3 = 0, \quad \alpha^4 = 200, \quad \alpha^5 = 180, \quad \alpha^6 = 200, \quad \alpha^7 = 0, \quad \alpha^8 = 0,$$
$$w_0 = -11{,}6.$$

(10 pts) (a) Calculez les valeurs du vecteur $\mathbf{w}$ de l'hyperplan séparateur de ce classifieur.

(10 pts) (b) Déterminez les données qui sont des vecteurs de support ainsi que les données qui sont dans la marge ou mal classées.

(10 pts) (c) Supposons maintenant que l'on veut traiter une donnée $\mathbf{x} = [0{,}37 \; 0{,}45]^\top$ avec ce SVM. Calculez la valeur $\mathrm{h}(\mathbf{x})$ correspondante (valeur réelle avant seuillage de la sortie).