Examen ING2 EILCO - Ingénierie Mathématique

Décembre 2022

Nom : Prénom :

Total: 27 points
Durée: 2h
Instructions générales: L'examen comprend 2 parties (Chacune de ces parties reprenant différentes sous-questions). Vous êtes libres de rédiger vos réponses sur des pages supplémentaires en veillant toutefois à bien indiquer le numéro de chaque question. Une fois l'examen terminé, Assurez vous de bien écrire votre nom (de façon lisible) sur chacune des pages. Répondez à un maximum de questions, en commençant par les questions qui vous semblent les plus abordables.

Question 1 (14pts)

1. [5*pts*] *Indiquer si les affirmations suivantes sont vraies ou fausses*

Vrai / Faux	*Minimiser la fonction d'entropie binaire croisée correspond à rechercher un estimateur de maximimum de vraissemblance*
Vrai / Faux	*Le modèles de régression logistique est un exemple de modèle de classification de type génératif*
Vrai / Faux	*Dans le cadre de la régularisation de type Ridge et pour un coefficient de régularisation $\lambda > 0$, augmenter la valeur de λ résulte en une augmentation de la variance de la famille de modèles correspondants*
Vrai / Faux	*Dans le cadre de la régression linéaire, ajouter une pénalité de type Ridge, de coefficient associé $\lambda > 0$, correspond à translater les valeurs propres de la matrice $\boldsymbol{X}^T\boldsymbol{X}$ de λ*
Vrai / Faux	*Les équations normales admettent toujours au moins une solution*
Vrai / Faux	*Dans le cas d'un modèle de classification linéaire, initialiser l'ensemble des poids et des biais à zéro aura pour conséquence de n'entrainer aucune mise à jour de ces paramètres dans le cadre d'une descente de gradient*
Vrai / Faux	*Une descente de gradient appliquée à un modèle linéaire et à une fonction de coût donnée par la somme des carrés des erreurs, retourne toujours le minimum global.*

2. [3*pts*] *Donner le pseudo-code pour le modèle de classification "un contre tous"*

3. [6*pts*] *On considère le modèle de régression suivant, connu sous le nom de "réseau élastique",*

$$L\left(\beta, \left\{x^{(i)}, t^{(i)}\right\}_{i=1}^{N}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(t^{(i)} - \beta_0 - \sum_{j=1}^{D}\beta_j x_j^{(i)}\right)^2 + \lambda_2\left(\sum_{j=1}^{D}|\beta_j|^2\right) + \lambda_1\left(\sum_{j=1}^{D}|\beta_j|\right) \tag{1}$$

   (a) [1*pt*] *Indiquer la partie dérivable et la partie non dérivable de la fonction coût.*

(b) [2pts] *On a représenté, à la Figure 2, l'évolution des coefficients de régression (chacun des β_j est représenté par une courbe) obtenus en minimisant la fonction de coût* (1) *pour différents choix de* (λ_1, λ_2). *En particulier, chacun des différents cas suivants est représenté par l'une des 3 figures:*

- *Le modèle Ridge* ($\lambda_1 = 0, \lambda_2 > 0$)
- *Le modèle LASSO* ($\lambda_2 = 0, \lambda_1 > 0$)
- *Un modèle intermédiaire, donné par des coefficients* $\lambda_1, \lambda_2 > 0$ *avec* $\lambda_1 = 9\lambda_2$

Indiquer, sur chacune des figures, le modèle correspondant.

(c) [3pts] *On considère le projecteur suivant dont la* $i^{\text{ème}}$ *composante est définie par*

$$[\mathcal{T}_{\lambda t}(\boldsymbol{\beta})]_i = \max(0, |\beta_i| - \lambda t)\operatorname{sign}(\beta_i) \tag{2}$$

Le projecteur remplace donc par zéro les coefficients β_i *dont l'amplitude est inférieure à* λt. *À l'aide de ce projecteur, donner un algorithme de minimisation de la fonction de coût* (1).

Question 2 (13pts)

1. [5pts] *Indiquer si les affirmations suivantes sont vraies ou fausses*

Vrai / Faux	*Augmenter le nombre de neurones dans la couche cachée d'un réseau à une couche augmente la variance de la famille de modèles correspondants*
Vrai / Faux	*Un réseau de neurones profond dont toutes les fonctions d'activation sont données par l'identité est équivalent à un modèle linéaire*
Vrai / Faux	*L'algorithme de descente de gradient appliqué à un réseau de neurones avec un taux d'apprentissage suffisamment faible et un nombre suffisamment grand d'itérations converge nécessairement vers le minimum global de la fonction coût*
Vrai / Faux	*Le nombre minimum de neurones nécessaires à l'apprentissage d'un modèle de type 'OU Exclusif' (XOR) est égal à 4*
Vrai / Faux	*En terme d'expressivité (i.e. capacité à capturer une distribution de données), un réseau de neurones est plus puissant qu'un modèle linéaire basé sur des caractéristiques polynomiales*

2. [3pts] *On considère le réseau de neurones représenté à la Figure 1, reprenant 3 couches cachées. Les poids associés à l'unité i de la couche ℓ sont représentés par la variable $w_{ij}^{(\ell)}$. Chaque neurone est représenté par une fonction d'activation sigmoïdale σ, et muni d'un biais $w_{i0}^{(\ell)}$ (non représenté sur la figure)*

   (a) [1pts] *Esquisser la fonction sigmoïde*

   (b) [2pts] *Donner l'expression détaillée de $y(\boldsymbol{x})$ en fonction de $\boldsymbol{x} = (x_1, x_2, x_3)$ et des paramètres $w_{ij}^{(\ell)}$.*

3. [5pts] *On souhaite utiliser l'algorithme de Backpropagation, afin de calculer le gradient de l'entropie binaire croisée (pour une paire $(\boldsymbol{x}^{(i)}, t^{(i)})$) par rapport au poids $w_{11}^{(0)}$ du réseau de neurones représenté à la Figure 1. Pour ce faire on procédera comme suit:*

   (a) [1pts] *Donner l'expression de la fonction d'entropie binaire croisée pour la paire $\left\{\boldsymbol{x}^{(i)}, t^{(i)}\right\}$*

   (b) [1pts] *Donner l'expression du $\delta^{(2)} = \delta_{out} = \frac{\partial L}{\partial a_{out}}$ (dérivée de la fonction d'entropie par rapport à la valeur de préactivation en sortie du réseau)*

   (c) [2pts] *Donner l'équation de backpropagation et utiliser cette équation pour déduire, à partir de δ_{out}, la valeurs des δ_i^1 pour $i = 1, 2$. Ensuite déduire, à partir des δ_i^1, la valeur de δ_1^0.*

   (d) [1pts] *Finalement, donner l'expression de la dérivée $\frac{\partial L}{\partial w_{11}^0}$ de la fonction de coût par rapport au poids w_{11}^0 en fonction de δ_1^0 et de $z_1^{(0)} = x_1$. En déduire, à partir de votre expression de δ_1^0 la réponse à la question.*

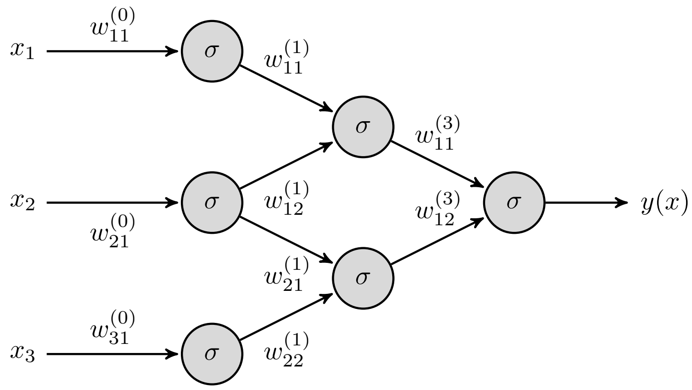

Figure 1: Réseau de neurones utilisé pour les questions 2.2 et 2.3

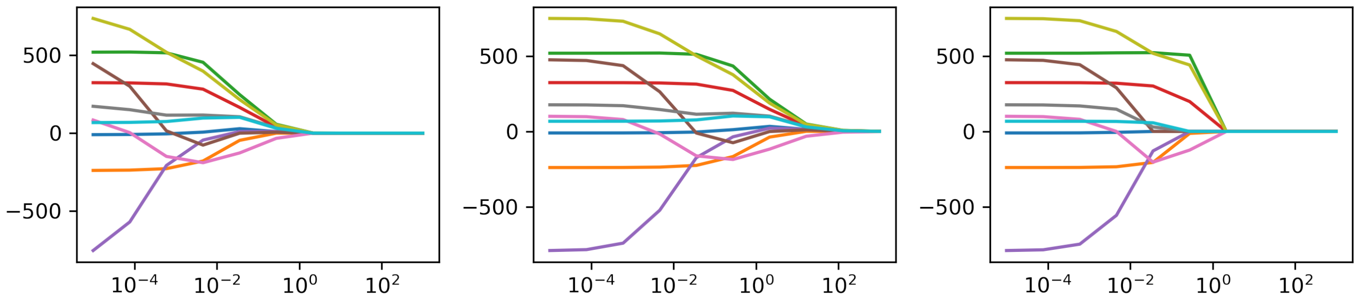

Figure 2: Évolution des coefficients de régression pour une valeur croissante des coefficients de régularisation λ_1, λ_2 dans le modèle de type "réseau élastique". Les différentes courbes correspondent aux différents coefficients.