

Examen ING2 EILCO - Ingénierie Mathématique

Mai 2023

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 (12pts)

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

Vrai / Faux *Le modèle de régression logistique est plus robuste (vis-a-vis des donnée aberrantes) que le modèle de régression linéaire*

Vrai / Faux *Rechercher la solution correspondant au maximum du logarithme d'une fonction positive est équivalent à rechercher la solution correspondant au maximum de la fonction elle-même*

Vrai / Faux *Pour un même ensemble de données, un modèle de type Ridge aura toujours d'avantage de coefficients nuls qu'un modèle de type LASSO*

Vrai / Faux *La fonction d'entropie binaire croisée correspond à l'opposé du logarithme de la fonction de vraisemblance pour un modèle de classification binaire dont les probabilités sont définies par la fonction sigmoïde*

Vrai / Faux *Soit le modèle $\frac{1}{N} \sum_{i=1}^N (t^{(i)} - h_{\beta}(\mathbf{x}^{(i)}))^2 + \lambda \sum_{j=1}^D |\beta_j|^2$. Augmenter la valeur de λ aura pour conséquence de réduire l'amplitude des coefficients β_j*

Vrai / Faux *Les régularisations de type Ridge et LASSO peuvent être réécrites à la fois sous la forme de problèmes d'optimisations non contraints et sous contraintes.*

2. [3pts] Expliquer la différence entre modèle génératif et discriminant. Donner un exemple de modèle pour chacune de ces deux familles.

3. [4pts] On se place dans le cadre de la régression linéaire. En particulier on s'intéresse à l'évolution des termes de biais et variance pour deux familles de modèles (correspondant à deux niveaux de régularisation). Les prédictions obtenues pour les deux familles, entraînées sur un même ensemble de données (indiqué en rouge) sont représentées à la Figure 2 (haut et bas).

(a) [2pts] Indiquer sur chacune des figures si la figure correspond à un biais élevé ou faible et à une variance élevée ou faible.

(b) [2pts] L'un des figures correspond à un niveau de régularisation $\lambda = 1e^8$ tandis que l'autre correspond à un niveau de régularisation $\lambda = 1e^{-4}$. Pour chaque valeur de λ , indiquer la figure correspondante.

Question 2 (15pts)

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

- Vrai / Faux La fonction d'entropie croisée catégorielle peut être utilisée dans le cadre d'un problème de classification comportant plus de deux classes
- Vrai / Faux Un réseau de neurones dont toutes les fonctions d'activation sont des fonctions de type Relu ($f(x) = \max(0, x)$) ne peut pas apprendre un ensemble de données de type "OU exclusif" (XOR)
- Vrai / Faux ADAPtive Moment estimation (ADAM) est un algorithme pouvant être utilisé lors de l'étape d'apprentissage dans les réseaux de neurones
- Vrai / Faux Entraîner un réseau de neurones à l'aide d'une descente de gradient stochastique est moins coûteux au niveau mémoire qu'entraîner un réseau de neurones à l'aide d'une descente de gradient sur l'ensemble des données
- Vrai / Faux Une descente de gradient stochastique sur un modèle de régression linéaire (avec fonction de coût donnée par la somme des carrés des résidus) converge toujours vers le minimum global de la fonction de coût reprenant l'ensemble des données.

2. [3pts] On souhaite implémenter un réseau convolutionnel pour classer différentes images en couleur.

- (a) [2pts] Lister les différents types de couches que vous envisageriez d'utiliser.
- (b) [1pts] Si la plus petite image est de taille $h \times w$, et si l'on ne souhaite pas extrapoler les données, quelles sont les dimensions des entrées de la première couche?

3. [2pts] 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) [2pts] Donner l'expression de la sortie du premier neurone de la troisième couche en fonction des valeurs de post-activation de la deuxième couche, des poids et biais, et de la fonction d'activation.

4. [5pts] On souhaite utiliser l'algorithme de Backpropagation, afin de calculer le gradient de l'entropie binaire croisée (pour une paire $(\mathbf{x}^{(i)}, t^{(i)})$) par rapport au poids $w_{12}^{(2)}$ de la couche 2 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 $\{\mathbf{x}^{(i)}, t^{(i)}\}$
- (b) [1pts] Donner l'expression du $\delta^{(4)} = \delta_{out} = \frac{\partial L}{\partial a_{out}}$ (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 $\delta_i^{(k)}$ pour $k = 3, 2$.
- (d) [1pts] Finalement, donner l'expression de la dérivée $\frac{\partial L}{\partial w_{12}^{(2)}}$ de la fonction de coût par rapport au poids $w_{12}^{(2)}$ en fonction de $\delta_1^{(2)}$ et de $z_2^{(1)}$.

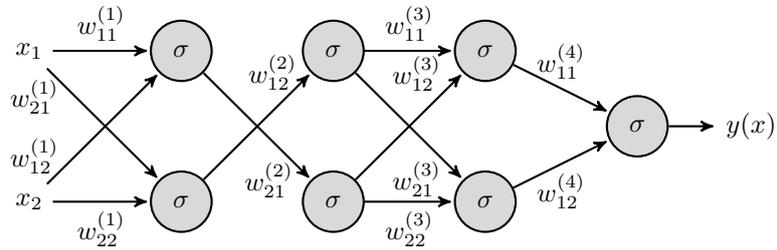


Figure 1: Réseau de neurones utilisé pour les questions 2.3 et 2.4

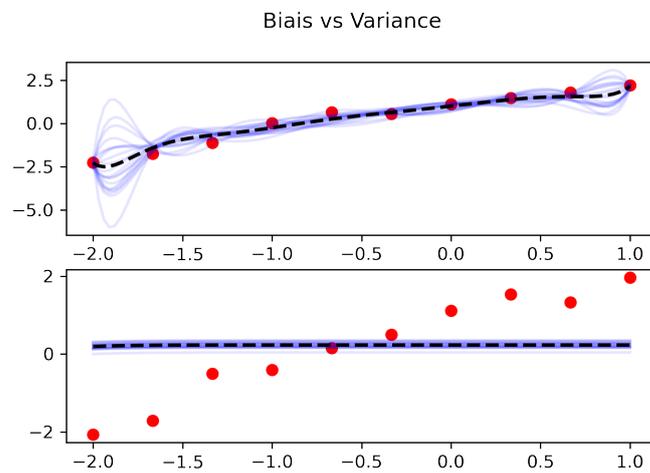


Figure 2: Figure utilisée à la question 1.3.