Réseau de neurone peu profond
- L2 Math, IMIA -

1. Considérons le modèle linéaire suivant

$$h(\mathbf{x}) = b + \mathbf{x}^T \mathbf{w}$$

où $\mathbf{w}$ et $\mathbf{x}$ sont des vecteurs de $\mathbb{R}^2$ et $\mathbf{w}^T$ est la transposée de $\mathbf{w}$.
On suppose que les paramètres de ce réseau de neurones (sans couche cachée) prennent les valeurs $b = 0, w_1 = 1, w_2 = .5$.

   (a) Représenter h sous la forme d'un graphe (réseau de neurones). Préciser la (ou les) fonction(s) d'activation.

   (b) Quelle sortie (ou réponse) va renvoyer le réseau de neurones pour $\mathbf{x} = \begin{bmatrix} 2 & 3 \end{bmatrix}^T$?

   (c) On dispose des observations suivantes E :

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ -2 & 8 \\ 0 & 1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 6 \\ 1 \end{bmatrix}$$

   Chaque ligne de $\mathbf{X}$ représente une observation et chaque entrée de $\mathbf{y}$ représente la réponse associée.

   (d) Calculer l'erreur

$$L(b, \mathbf{w}) = \frac{1}{2} \sum_{e \in E} (y_e - b - \mathbf{x}_e^T \mathbf{w})^2$$

   pour cet ensemble de données.

   (e) Peut-on trouver des poids qui conduisent à une erreur plus faible ? Que proposez vous ? Mettre en oeuvre (au moins une partie) de la méthode proposée.

2. On considère la fonction g définie par

$$\begin{aligned} g(x) &= x \text{ si } x \in [0, 1] \\ g(x) &= 2 - x \text{ si } x \in [1, 2] \\ g(x) &= 0 \text{ sinon} \end{aligned}$$

   (a) Représenter la fonction g sur $[-2, 4]$.

   (b) Montrer qu'on peut écrire la fonction g sous la forme

$$\begin{aligned} \Phi(x) &= \max(w_{21} \max(w_{11}x + b_{11}, 0) + w_{22} \max(w_{12}x + b_{12}, 0) + b_2, 0) \\ &= (w_{21}(w_{11}x + b_{11})_+ + w_{22}(w_{12}x + b_{12})_+ + b_2)_+ \end{aligned}$$

   où $z_+ = z$ si $z > 0$ et 0 sinon.
   Puis identifier les valeurs des poids et des biais.

   (c) Représenter Φ sous la forme d'un graphe (réseau de neurones).

(d) Quelles sont les fonctions d'activation de la couche cachée et de la couche de sortie ?

(e) Dans la suite on note $\mathbf{w}$ le vecteur contenant les poids et les biais.
Soit un couple d'observations (x, y) tel que $x = 1/2$, $y = 1$. Calculer le gradient de la fonction de perte des moindres carrés $L(\mathbf{w}) = \frac{1}{2}(y - \Phi(\mathbf{w}, x))^2$ par rétro-propagation. On pourra calculer seulement certaines composantes :

$$\frac{\partial L(\mathbf{w})}{\partial w_{22}}, \frac{\partial L(\mathbf{w})}{\partial w_{12}}, \frac{\partial L(\mathbf{w})}{\partial b_{12}}$$

(f) Que peut-on faire si on veut modifier le poids w_{22} de façon à ce que le réseau de neurones donne une sortie plus proche de $y = 1$ pour l'entrée $x = 1/2$? Mettre en oeuvre (au moins une partie) de la méthode proposée.