EXAMEN | MOBI | 27/03/2019 $08^{30}$–$11^{30}$

**aucun document – aucune machine – une simple calculatrice**

cinq exercices indépendants – trois questions de théorie

*Lorsque des calculs sont nécessaires, il est impératif de les présenter sur la feuille d'examen. La qualité de la rédaction et la clarté de la présentation seront prises en compte dans la notation : les réponses devront être* ***précises*** *et* ***argumentées****.*

# Exercice 1

Définir un perceptron qui calcule la fonction booléenne de trois variables, définie par la table de vérité suivant.

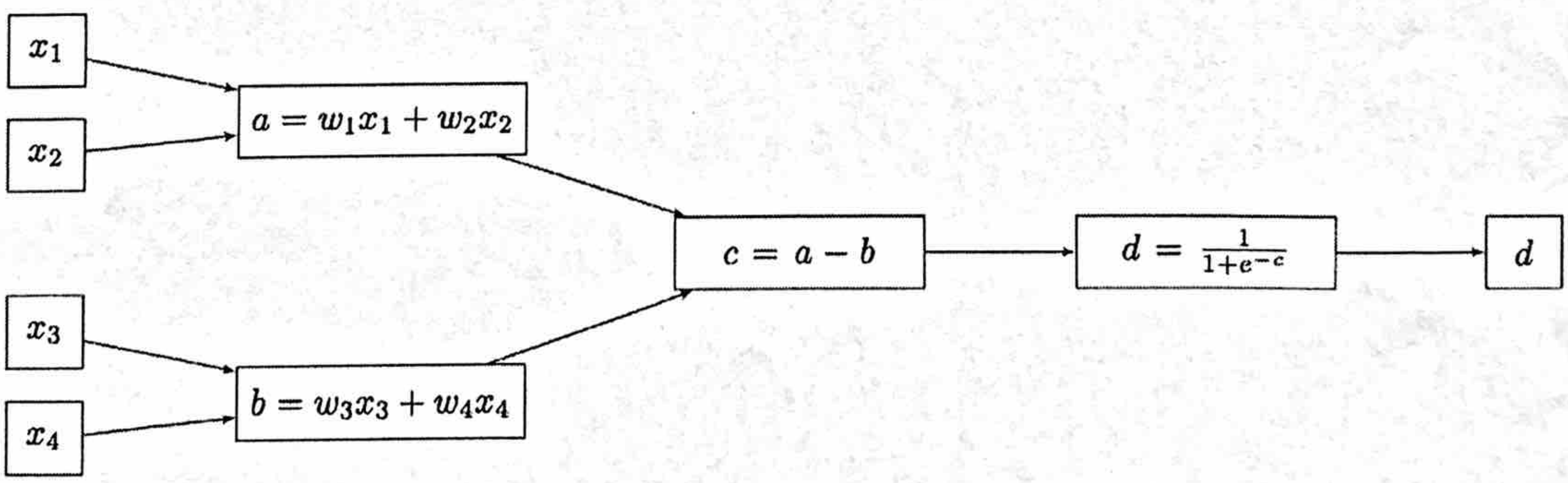
| $x_1$ | $x_2$ | $x_3$ | $y$ |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 |
| 0 | 1 | 1 | 1 |
| 1 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 |
| 1 | 1 | 0 | 1 |
| 1 | 1 | 1 | 0 |

*Théorie.* Montrer que si on modifie la table de telle sorte que la sortie $y$ correspondant à $(0,0,0)$ devienne 1, il n'y a pas de perceptron capable de calculer la fonction.

(1, 0, 1)

# Exercice 2

Considérons le réseau suivant.




Avec les entrées $x_i = i$, le réseau devrait produire la sortie $y = 1$. Les poids initiaux sont $w_1 = 1$, $w_2 = -1$, $w_3 = 1$, et $w_4 = -1$. Quelle est la valeur produite par le réseau ? Si la fonction de perte est $L = \frac{1}{2}(d-y)^2$ et si le taux d'apprentissage est $\eta = 1$, comment les poids changent-ils après une application de la rétropropagation ? Quelle est la nouvelle valeur produite par le réseau ?

# Exercice 3

Considérons un réseau de neurones avec trois neurones d'entrée, trois neurones de sortie et une couche cachée avec deux neurones. La fonction d'activation de tous les neurones est la fonction sigmoïde. Supposons que $w^2 = \begin{pmatrix} -2 & 2 & -2 \\ 1 & 1 & -1 \end{pmatrix}$ et $w^3 = \begin{pmatrix} 1 & -3.5 \\ 0.5 & -1.2 \\ 0.3 & 0.6 \end{pmatrix}$, et que tous les biais sont égaux à 0. Calculer les sorties du réseau avec les entrées 2, 3, et 1. Si les sorties désirées sont 0.5, 0.7, et 0.2 et le taux d'apprentissage est $\eta = 1$, calculer la nouveau valeur du poids $w^2_{2,3}$ après une application de la rétropropagation (rappelez-vous que les quatre équations de rétropropagation sont : (1) $\delta^L = \nabla_a C \otimes \sigma'(z^L)$, (2) $\delta^l = \left( \left(w^{l+1}\right)^T \delta^{l+1} \right) \otimes \sigma'(z^l)$, (3) $\frac{\partial C}{\partial b^l_j} = \delta^l_j$, et (4) $\frac{\partial C}{\partial w^l_{j,k}} = \delta^l_j a^{l-1}_k$).

# Exercice 4

Considérons un réseau de neurones avec les couches suivantes.

- Input : image $227 \times 227$ avec 3 canaux
- Couche convolutive avec 3 filtres, chacun de taille $11 \times 11 \times 3$, avec décalage de 4 et sans padding.
- Max pooling de taille $3 \times 3$, avec décalage de 2
- Couche entièrement connectée avec 10 sorties.

Calculer la taille et le nombre de paramètres de chaque niveau du réseau.

*Théorie.* Quel type de filtre utiliserions-nous pour détecter les bordures horizontales dans une image en noir et blanc ? Et pour les bordures verticales ?

# Exercice 5

Définir un réseau de neurones récurrent capable de calculer le signe de la somme d'une séquence de nombres entiers dans l'ensemble $\{-1, 1\}$. Plus précisément, étant donné une séquence $x_t$ de nombres entiers dans l'ensemble $\{-1, 1\}$, le réseau doit calculer, à chaque instant $t$, la valeur 1 si $x_t + s_{t-1} \geq 0$, où $s_{t-1}$ est la somme à l'instant $t-1$, sinon la valeur 0. Par exemple, si la séquence commence avec

$$1, 1, -1, 1, -1, -1, -1, 1, -1, 1, 1, -1,$$

alors la sortie du réseau doit être

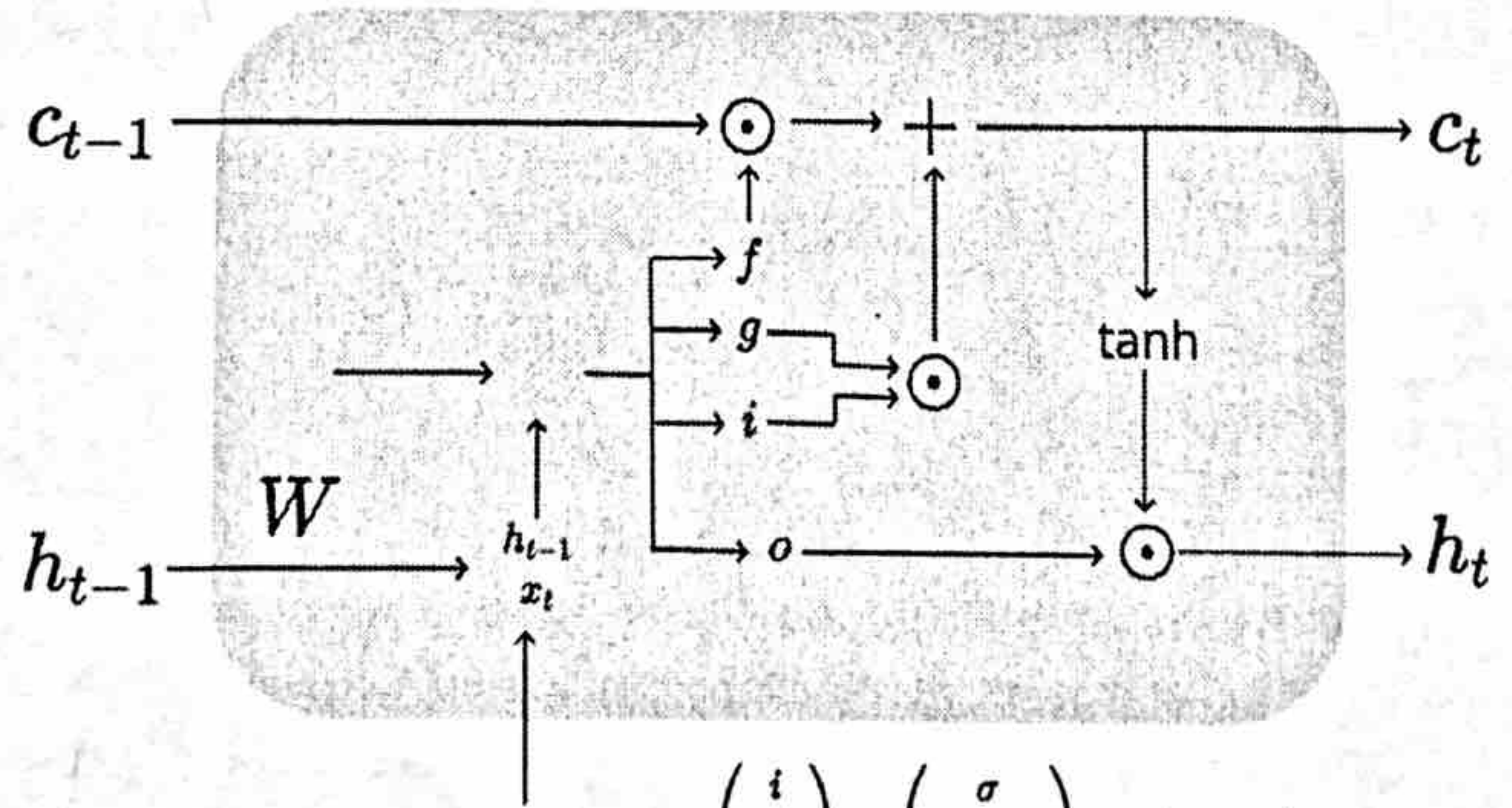
$$1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1,$$

puisque les valeurs de la somme sont

$$1, 2, 1, 2, 1, 0, -1, 0, -1, 0, 1, 0.$$

Pour définir le réseau, vous devez spécifier : (1) combien de neurones cachés sont dans le réseau, (2) combien de neurones de sortie sont dans le réseau, (3) la fonction d'activation des neurones cachés et des neurones de sortie, et (4) les poids des connexions et les biais.

*Théorie.* Expliquer en quelques mots quelle est la signification des fonctions $f$, $g$, $i$, et $o$ utilisées dans un réseau de neurones LSTM, comme indiqué dans la figure suivante.




$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} W \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}$$

$$c_t = f \odot c_{t-1} + g \odot i \qquad h_t = o \odot \tanh(c_t)$$

Expliquer également pourquoi ce type de réseau ne souffre pas du phénomène de disparition du gradient.