

Réseaux neuronaux  
IFT 780

## Examen intra

Professeur : Pierre-Marc Jodoin

---

Signature :Cip :

---

**Consignes**

1. Cet examen comporte **cinq (5) questions** au total.
  2. Vous avez droit à des notes manuscrites non-photocopiées.
  3. Vous avez droit à une calculatrice sans capacité de communication.
- 

**Question 1 (20 points)**

Avant d'entraîner un réseau de neurones, il faut toujours en initialiser les paramètres. Soit un réseau de neurones multicouche pleinement connecté comprenant 5 couches cachées et dont l'objectif est de classifier des images CIFAR10 en 10 classes. Le réseau utilise ReLU comme fonction d'action et une entropie croisée comme fonction de coût. Dites ce qui arrivera à l'entraînement ;

1. Si vous initialisez tous les poids du réseau à 1.
2. Si vous initialisez tous les poids du réseau à -1.
3. Si vous initialisez les poids du réseau avec différentes valeurs aléatoires contenues entre -0.01 et +0.01.
4. Si vous initialisez les poids du réseau avec différentes valeurs aléatoires contenues entre -100 et +100.

Justifiez mathématiquement vos réponses.

## Question 2 (20 points)

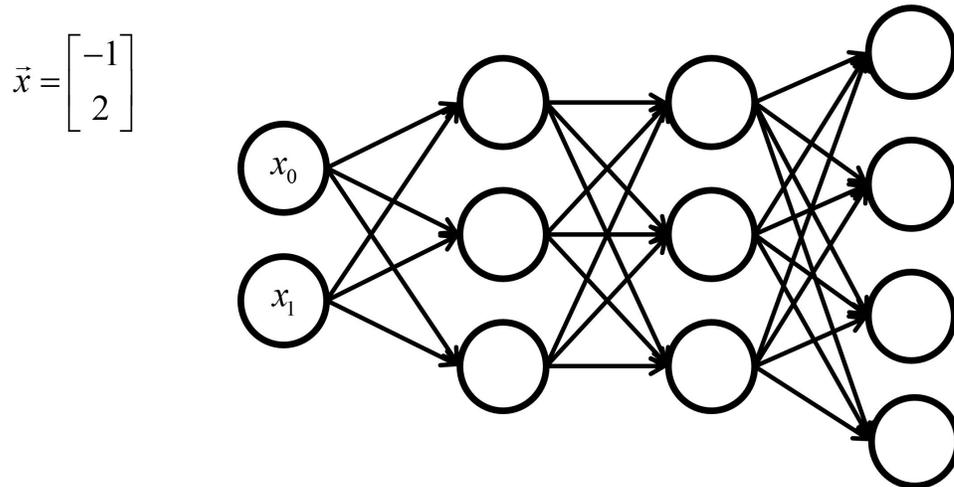
On vous demande le design d'un réseau à convolution contenant une centaine de couches. Dites quelles seraient les trois principales caractéristiques du design de ce réseau afin que ce dernier s'entraîne comme il se doit. Justifiez bien votre réponse.

### Question 3 (20 points)

Soit le réseau de neurones de la figure 1 prenant en entrée un vecteur à deux dimensions et dont l'objectif est de faire la régression de quatre (4) variables. La fonction de coût pour une paire  $(\vec{x}_i, \vec{t}_i)$  est la fonction L2 suivante :

$$E(W) = \|\mathbf{y}_W(\vec{x}_i) - \vec{t}_i\|^2 \quad (1)$$

où  $y_W(\vec{x}_i)$  est la sortie du réseau et  $\vec{t}_i$  est le vecteur cible. En supposant que la fonction d'activation ReLU est utilisée, veuillez (a) donner la sortie de chaque neurone suite à une propagation avant (à inscrire dans les cercles de la figure 1) et (b) donner le gradient  $\nabla_{W^{[2]}} E(W)$  pour la cible  $\vec{t}_i = (0, 10, 20, 3)^T$ .



$$W^{[0]} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ -1 & 0.5 \end{bmatrix} \quad W^{[1]} = \begin{bmatrix} 0 & -1 & 0 \\ 2.25 & 2 & -3 \\ 0.13 & 3 & 1 \end{bmatrix} \quad W^{[2]} = \begin{bmatrix} 3 & 0.5 & -1 \\ 3 & 3 & 1 \\ -7.1 & 1 & 2 \\ 0 & -0.5 & -0.1 \end{bmatrix}$$

FIGURE 1 – Figure associée à la question 3.

## Question 4 (20 points)

Un médecin sollicite votre aide pour développer un logiciel d'aide au diagnostique. Le but de ce logiciel est de distinguer cinq(5) sortes d'épilepsie, et ce à l'aide de signaux 1D issus d'un électroencéphalogramme (EEG).

1. Croyez-vous qu'un réseau à convolution à  $N$  couches serait plus ou moins performant qu'un réseau pleinement connectés comprenant également  $N$  couches et ce, dans le cas où les deux réseaux seraient entraînés avec la même fonction de coût et les mêmes données ?
2. Nonobstant votre réponse à la question précédente, donnez la structure d'un réseau à convolution pleinement convolutif viable de votre cru comprenant 5 couches au total. Veuillez préciser toutes les informations associés aux couches de votre réseau et considérez qu'une entropie croisée est utilisée. Votre réseau doit également pouvoir s'accommoder de signaux de taille allant de 300 à 600 échantillons (la figure 2 illustre un signal ayant 300 échantillons). Votre réponse doit prendre la forme d'un schéma comme celui de la figure 3.

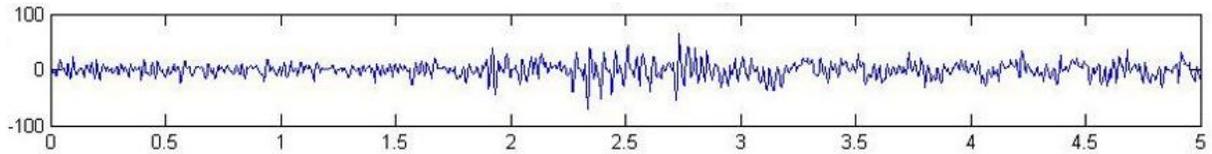


FIGURE 2 – Figure associée à la question 3 illustrant un signal EEG d'une durée de 5 secondes et comprenant 60 échantillons par seconde.

## Question 5 (20 points)

Donnez pour chaque couche du réseau convolutif de la figure 3 le nombre de paramètres devant être appris par rétro-propagation ainsi que le nombre de neurones.

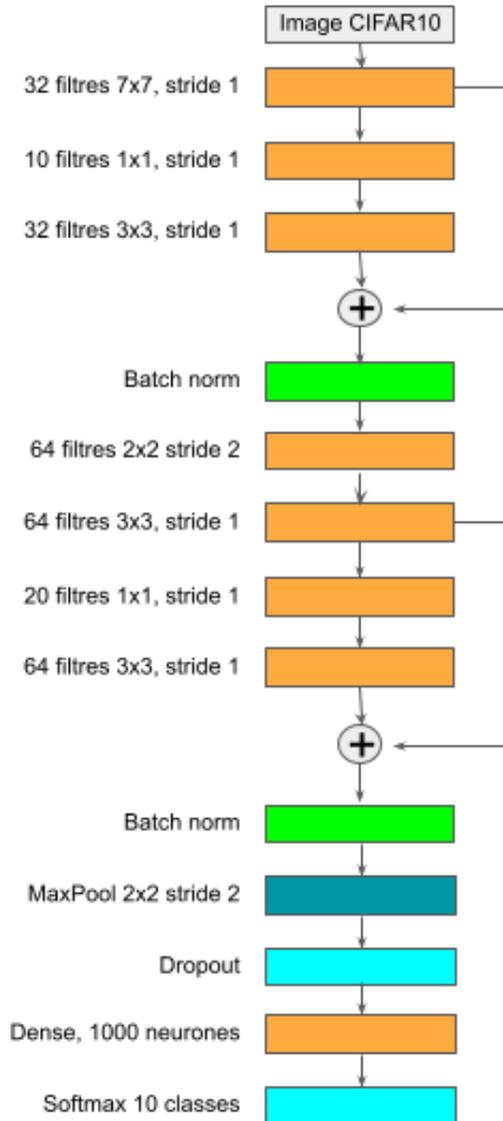


FIGURE 3 – Figure associée à la question 4.

Fin de l'examen.