Examen final - 5 février 2021 - Durée : 1 h

Document interdit, téléphone interdit, calculatrice interdite.

Les réponses doivent être ***détaillées****.*

Nom : Prénom :

Ex. 1 :	Ex. 2 :	Ex. 3 :	Ex. 4 :	Ex. 5 :	Total (/20) :

Exercice 1 (Encodage, ≃ 3 pts)

Considérons une étiquette y qui peut prendre K valeurs discrètes $1, 2, \dots, K$.

1. Expliquer comment fonctionne le "one-hot encoding" appliqué à y.

2. On suppose qu'un réseau de neurones cherche à prédire y. La dernière couche du réseau de neurones est une couche pleinement connectée ("fully connected"). Combien de neurones doit comporter cette dernière couche ? Quelle fonction faut-il appliquer à la sortie, notée h, de cette couche pour obtenir une sortie $\hat{y}$ comparable avec y ?

3. Décrire la fonction de perte "cross-entropy" $L(\hat{y}, y)$ employée pour mesurer la qualité de la sortie $\hat{y}$ par rapport à une étiquette y encodée avec le "one-hot encoding".

Exercice 2 (Approximation ReLU, ≃ 5 pts)

1. Dessiner la fonction $f(x)$ approchée par le réseau ReLU à une couche cachée suivant

$$f(x) = \mathrm{ReLU}\big(\mathrm{ReLU}(x) - \mathrm{ReLU}(2x-1)\big).$$

2. Soit $g(x)$ la fonction linéaire par morceau définie par

$$\begin{cases} x & \text{if } x \in [0,1], \\ 2-x & \text{if } x \in [1,2], \\ 0 & \text{sinon.} \end{cases}$$

Contruire un réseau de neurones avec 1 couche cachée et la fonction d'activation ReLU pour approximer la fonction $g(x)$.

Exercice 3 (Disparition du gradient, $\simeq$ 5 pts)
On considère le réseau

$$f(x) = \sigma(w_3\sigma(w_2\sigma(w_1 x)))$$

où $x \in \mathbb{R}$ est l'entrée du réseau et $w_1,\ w_2,\ w_3 \in \mathbb{R}$ sont les poids du réseau. La fonction $\sigma(\cdot)$ désigne la fonction sigmoïde.

1. En utilisant le formalisme de l'algorithme de rétropropagation du gradient, calculer $\frac{\partial f}{\partial w_1}(x)$.

2. On suppose que les poids w_1, w_2, w_3 vérifient $-1 \le w_i \le 1$ pour $i = 1, 2, 3$. Montrer que

$$\left|\frac{\partial f}{\partial w_1}(x)\right| \le \left(\frac{1}{4}\right)^{\frac{1}{3}} |x|.$$

3. Expliquer à partir de la question précédente ce qu'on appelle le problème de la disparition du gradient ("vanishing gradient").

Exercice 4 (Fast-RCNN, $\simeq$ 4 pts)
Décrire, à l'aide d'un schéma, l'architecture du détecteur d'objet Fast-RCNN. On ne vous demande pas de donner tous les détails de l'architecture mais les éléments principaux doivent être clairement mis en avant et expliqués.

Exercice 5 (Transformer, $\simeq$ 3 pts)
On considère une séquence d'entrée $x_1, x_2, \ldots, x_t$ où chaque $x_i \in \mathbb{R}^d$.

Nous souhaitons calculer la séquence $y_1, y_2, \ldots, y_t$ en sortie d'un modèle d'auto-attention ("selft-attention") d'un transformer où chaque $y_i \in \mathbb{R}^d$.

Donner l'équation qui relie la sortie y_i aux entrées $x_1, x_2, \ldots, x_t$. Bien définir et expliquer les différents éléments de cette équation.

Signature :