Prélaboratoire : L'apprentissage par réseau de neurones - Partie 2

Nom : ______________________ Groupe : ____________

Les directives du laboratoire sont disponibles sur la page web de ce projet :

https://physique.cmaisonneuve.qc.ca/svezina/projet/apprentissage_reseau/apprentissage_reseau.html

Mise en situation 2

Lors de la mise en situation 1, vous avez réalisé le calcul

$$a = \tanh(Wx + b)$$

où $a = \tanh(z)$ tel que $z = Wx + b$ avec la matrice W et les vecteurs x et b suivant :

$$W = \begin{pmatrix} 1.0 & 4.0 \\ 3.0 & 2.0 \\ -2.0 & -3.0 \end{pmatrix} \quad , \quad x = \begin{pmatrix} -0.4 \\ 0.6 \end{pmatrix} \quad \text{et} \quad b = \begin{pmatrix} 0.8 \\ 0.5 \\ -0.2 \end{pmatrix}$$

Dans le calcul de $a = \tanh(Wx + b)$, vous avez transformé un vecteur x de 2 composantes à l'aide d'un paramètre W et b en un vecteur a de 3 composantes.

Maintenant, nous allons <u>propager une erreur</u> Δ grâce à l'équation de la dérivée en chaîne

$$\frac{\partial C}{\partial x} = \Delta \frac{\partial a}{\partial z} \frac{\partial z}{\partial x} \quad \text{avec la donnée } \Delta = \begin{pmatrix} 0.2 \\ 0.1 \\ -0.1 \end{pmatrix}$$

que nous allons calculer grâce à l'expression[1]

$$\frac{\partial C}{\partial x} = \frac{\partial}{\partial x}\left((Wx + b) \circ \frac{\partial \tanh(z)}{\partial z} \circ \Delta \right) \quad .$$

(séquence de multiplication de droite à gauche)

Vous remarquerez que l'ordre des opérations mathématique seront réalisées en <u>séquence de droite à gauche</u> en raison du comportement non commutatif des opérations matricielles. Ce calcul transformera un vecteur Δ de 3 composantes en un vecteur final $\partial C / \partial x$ de 2 composantes.

Question 1.

La dérivée de la fonction tangente hyperbolique est égale à l'expression

$$\tanh'(z) = \frac{\mathrm{d} \tanh(z)}{\mathrm{d}z} = 1 - \tanh^2(z) \quad .$$

À partir des valeurs de votre vecteur colonne z obtenue à la question 2, vous devrez évaluer l'expression

$$\frac{\partial \tanh(z)}{\partial z} = 1 - \tanh^2(z)$$

pour l'ensemble des trois composantes du vecteur.

[1] L'opérateur matriciel $\circ$ correspond au produit de Hadamard et il vous sera présenté prochainement.

Prélaboratoire : L'apprentissage par réseau de neurones - Partie 2

Ainsi, pour évaluer $\partial a / \partial z$, vous devrez calculer ceci : (compléter l'espace vide)

$$\frac{\partial a}{\partial z} = \frac{\partial \tanh(z)}{\partial z} = \begin{pmatrix} \dfrac{\partial \tanh(z_0)}{\partial z} \\ \dfrac{\partial \tanh(z_1)}{\partial z} \\ \dfrac{\partial \tanh(z_2)}{\partial z} \end{pmatrix} = \begin{pmatrix} 1-\tanh^2(z_0) \\ 1-\tanh^2(z_1) \\ 1-\tanh^2(z_2) \end{pmatrix} = \begin{pmatrix} \quad \\ \quad \\ \quad \end{pmatrix}$$

(indiquez votre réponse ici : vecteur à 3 éléments)

Question 2.

Pour évaluer $\dfrac{\partial C}{\partial z} = \Delta \dfrac{\partial a}{\partial z}$, vous devrez effectuer le produit de Hadamard[2] $A \circ B$ (produit composante par composante) de deux vecteurs qui vous donnera l'expression suivante à calculer : (compléter l'espace vide)

$$\frac{\partial C}{\partial z} = \frac{\partial \tanh(z)}{\partial z} \circ \Delta = \begin{pmatrix} \dfrac{\partial \tanh(z_0)}{\partial z} \\ \dfrac{\partial \tanh(z_1)}{\partial z} \\ \dfrac{\partial \tanh(z_2)}{\partial z} \end{pmatrix} \circ \begin{pmatrix} \Delta_0 \\ \Delta_1 \\ \Delta_2 \end{pmatrix} = \begin{pmatrix} \dfrac{\partial \tanh(z_0)}{\partial z} \Delta_0 \\ \dfrac{\partial \tanh(z_1)}{\partial z} \Delta_1 \\ \dfrac{\partial \tanh(z_2)}{\partial z} \Delta_2 \end{pmatrix} = \begin{pmatrix} \quad \\ \quad \\ \quad \end{pmatrix}$$

(indiquez votre réponse ici : vecteur à 3 éléments)

Puisque le calcul réalisé correspond à l'erreur propagée par la fonction d'activation, nous pouvons également la nommer de façon plus compacte $\partial C / \partial z = \Delta^a$.

Question 6.

Lors de la mise en situation 1, vous avez réalisé le calcul

$$z = Wx + b = \begin{pmatrix} w_{00} & w_{01} \\ w_{10} & w_{21} \\ w_{20} & w_{21} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} + \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} w_{00}x_0 + w_{01}x_1 \\ w_{10}x_0 + w_{21}x_1 \\ w_{20}x_0 + w_{21}x_1 \end{pmatrix} + \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}$$

ce qui correspondait à la fonction d'agrégation z.

[2] Voir : https://fr.wikipedia.org/wiki/Produit_matriciel_de_Hadamard

Prélaboratoire : L'apprentissage par réseau de neurones - Partie 2

Maintenant, pour évaluer

$$\frac{\partial C}{\partial x} = \frac{\partial}{\partial x}\left(\left(Wx + b \right) \circ \frac{\partial \tanh\left(z \right)}{\partial z} \circ \Delta \right) = \frac{\partial}{\partial x}\left(\left(Wx + b \right) \circ \Delta^a \right) \quad \text{où} \quad \Delta^a = \frac{\partial \tanh\left(z \right)}{\partial z} \circ \Delta ,$$

vous devrez réaliser les calculs suivants en séquence :

1) Réaliser le produit de Hadamard $A \circ B$ entre $Wx + b$ et Δ^a tout en gardant les paramètres x et b sous leur forme algébrique, car ceux-ci seront affectés par une opération de dérivée. Ceci vous donnera une matrice 3 × 1.
2) Réaliser le produit matriciel entre le vecteur $\partial / \partial x$ (matrice 2 × 3) et le résultat précédent (matrice 3 × 1). Ceci vous donnera une matrice 2 × 1.
3) Effectuer l'opération de la dérivée $\partial / \partial x$ sur les paramètres x_0 et x_1.

Ainsi, pour évaluer $\partial C / \partial x$, vous devrez calculer ceci : (compléter les espaces vides en détaillant vos étapes de calcul)

$$\frac{\partial C}{\partial x} = \frac{\partial}{\partial x}\left(\left(Wx + b \right) \circ \Delta^a \right)$$

$$= \frac{\partial}{\partial x}\left(\left[\begin{pmatrix} w_{00}x_0 + w_{01}x_1 \\ w_{10}x_0 + w_{11}x_1 \\ w_{20}x_0 + w_{21}x_1 \end{pmatrix} + \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} \right] \circ \begin{pmatrix} \Delta_0^a \\ \Delta_1^a \\ \Delta_2^a \end{pmatrix} \right)$$

$$= \frac{\partial}{\partial x}\left(\begin{pmatrix} w_{00}x_0 + w_{01}x_1 \\ w_{10}x_0 + w_{11}x_1 \\ w_{20}x_0 + w_{21}x_1 \end{pmatrix} \circ \begin{pmatrix} \Delta_0^a \\ \Delta_1^a \\ \Delta_2^a \end{pmatrix} + \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} \circ \begin{pmatrix} \Delta_0^a \\ \Delta_1^a \\ \Delta_2^a \end{pmatrix} \right)$$

$$= \begin{pmatrix} \partial / \partial x_0 & \partial / \partial x_0 & \partial / \partial x_0 \\ \partial / \partial x_1 & \partial / \partial x_1 & \partial / \partial x_1 \end{pmatrix} \left(\begin{pmatrix} \Delta_0^a \left(w_{00}x_0 + w_{01}x_1 \right) \\ \Delta_1^a \left(w_{10}x_0 + w_{11}x_1 \right) \\ \Delta_2^a \left(w_{20}x_0 + w_{21}x_1 \right) \end{pmatrix} + \begin{pmatrix} \Delta_0^a b_0 \\ \Delta_1^a b_1 \\ \Delta_2^a b_2 \end{pmatrix} \right)$$

$$= \begin{pmatrix} \partial / \partial x_0 & \partial / \partial x_0 & \partial / \partial x_0 \\ \partial / \partial x_1 & \partial / \partial x_1 & \partial / \partial x_1 \end{pmatrix} \begin{pmatrix} \Delta_0^a \left(w_{00}x_0 + w_{01}x_1 \right) \\ \Delta_1^a \left(w_{10}x_0 + w_{11}x_1 \right) \\ \Delta_2^a \left(w_{20}x_0 + w_{21}x_1 \right) \end{pmatrix} + \begin{pmatrix} \partial / \partial x_0 & \partial / \partial x_0 & \partial / \partial x_0 \\ \partial / \partial x_1 & \partial / \partial x_1 & \partial / \partial x_1 \end{pmatrix} \begin{pmatrix} \Delta_0^a b_0 \\ \Delta_1^a b_1 \\ \Delta_2^a b_2 \end{pmatrix}$$

$$= \begin{pmatrix} \quad \\ \quad \end{pmatrix} + \begin{pmatrix} \quad \\ \quad \end{pmatrix}$$

$$= \begin{pmatrix} \quad \\ \quad \end{pmatrix}$$

$$= \begin{pmatrix} \quad \\ \quad \end{pmatrix}$$