

L'apprentissage par réseau de neurones – Partie 2

Questions conceptuelles

Question 7.1 :

Lors de l'activation de la fonction sigmoïde avec une grande valeur positive d'entrée, décrivez en mots la valeur que prendra la dérivée de cette fonction correspondant à cette valeur d'entrée.

Question 7.2 :

Qu'est-ce qu'on en commun la valeur des pentes des fonctions d'activation sigmoïde et tangente hyperbolique pour une valeur d'entrée égale à zéro ?

Question 7.4 :

Pourquoi la fonction d'erreur quadratique $C_{quad} = \frac{1}{2}(a_u^{(L)} - y_u)^2$ est mathématiquement plus intéressant qu'une fonction d'erreur linéaire de type $C_{lin} = a_u^{(L)} - y_u$?

Question 8.3 :

Dans l'équation de la propagation de l'erreur de la fonction d'agrégation $z_u^{(k)} = z_u^{(k)}(a_v^{(k-1)}, w_{uv}^{(k)}, b_u^{(k)})$, pourquoi retrouve t-on une sommation ce qui n'a pas été le cas pour la fonction d'activation $a_u^{(k)} = a_u^{(k)}(z_u^{(k)})$?

Question 8.6.1 :

Quel est le résultat de l'erreur propagée de votre réseau ?

Question 8.6.2 :

Est-ce que la l'erreur propagée de votre réseau sur JAVA correspond à la réponse de votre prélaboratoire ?

Question 9.3 :

Quel est le problème principal des valeurs du gradient de la fonction d'erreur lorsqu'on applique cette technique à un réseau contenant beaucoup de couches ?

Question 10.1.1 :

Pourquoi un terme négatif a été introduit dans « l'équation de l'apprentissage » ?

Question 10.1.2 :

Quel problème peut survenir lors de la phase de l'apprentissage si des erreurs ont été propagées par des fonctions d'activation où la pente était près de la valeur zéro ?

Question 11.1 :

Typiquement, un réseau non calibré est initialisé avec des valeurs aléatoires entre -1 et $+1$ pour l'ensemble des poids et biais que le constitue. Pourquoi ne pas tout simplement initialiser le tout avec des valeurs égales à zéro ?

Question 11.2.1 :

Avec le protocole d'entraînement tel que $CYCLE = 50$ et $LEARNING_RATE = 0.1$, est ce que vous avez suffisamment de vecteurs d'entraînement pour obtenir un taux de classification raisonnable ? Justifiez votre réponse.

Question 11.2.2 :

Est-il possible d'atteindre un taux de classification de plus de 95% ? Si oui, combien de vecteur d'entraînement ont été nécessaire et à quel taux d'apprentissage ?

Question 11.2.3 :

Pour obtenir un taux de classification supérieur à 95%, est-il nécessaire de réaliser un entraînement sur l'ensemble des vecteurs d'entraînement (683 vecteurs disponibles) ? Justifiez votre réponse.

Question 11.2.4 :

Est-il possible d'utiliser moins de 50 vecteurs d'entraînement et obtenir un taux de classification de plus de 90% ? Si oui, quel taux d'apprentissage ($LEARNING_RATE$) avez-vous utilisé ?

Question 11.3.1 :

Avec le protocole d'entraînement tel que $CYCLE = 100$, $ITERATION = 100$ et $LEARNING_RATE = 0.1$, est ce que vous avez suffisamment de vecteurs d'entraînement (10 000 vecteurs) pour obtenir un taux de classification raisonnable ? Justifiez votre réponse.

Question 11.3.2 :

Selon vous, avec un protocole d'entraînement tel que $LEARNING_RATE = 0.1$,

- a) Quel est le meilleur taux de classification que le réseau $784 \times 16 \times 16 \times 10$ peut obtenir ? Modifiez au besoin les paramètres $CYCLE$ et $ITERATION$ pour faire varier votre protocole d'entraînement.
- b) Combien de vecteur d'entraînement ont été nécessaire pour obtenir ce meilleur taux de classification ?

Question 11.3.3 :

Avec le protocole d'entraînement $CYCLE = 100$, $ITERATION = 100$, et $LEARNING_RATE = 0.9$, le taux de classification du réseau évolue tout au long de l'entraînement de façon particulier. Identifiez

- a) Une caractéristique positive.
- b) Une caractéristique négative.