

# Fiche d'auto-évaluation 04

January 4, 2022

Caractéristique = Feature

1. Quelle est la différence entre un réseau de neurones profond (deep) et non (shallow)?
2. Que veut-dire forward propagation? Et backpropagation (rétro-propagation)?
3. Un réseau de neurones peut approximer n'importe quelle fonction. Est-il correct?
4. Expliquez précisément quelle est la différence entre un réseau de neurones utilisé pour faire de la régression et un réseau de neurones utilisé pour la classification.
5. Écrivez l'équation qui décrit la sortie d'un neurone  $q$  dans la couche  $l$ .
6. La sortie d'un réseau de neurone est-elle linéaire ou non linéaire, par rapport à l'entrée?
7. Comment les non-linéarités sont introduites dans un réseau de neurones?
8. La rétro-propagation (backpropagation), est-elle utilisée pendant l'entraînement ou la prédiction?
9. Peut-on définir la régression logistique comme un réseau de neurones? Si oui, dans quel sens?
10. Écrivez la fonction de perte d'un réseau de neurones
11. Écrivez l'équation de descente de gradient pour minimiser une fonction  $u(\mathbf{z})$
12. Pour quelle raison on utilise la descente de gradient dans les réseaux de neurones?
13. Quand on applique la descente de gradient à un réseau de neurones, de quelle fonction calcule-t-on le gradient? Par rapport à quelles variables calcule-t-on le gradient?

14. A-t-on la garantie que, en appliquant la descente de gradient, on converge vers le minimum de la fonction de perte dans le cas de réseau de neurones? Et dans le cas de régression logistique? Expliquez pourquoi.
15. Quelle est la relation entre descente de gradient et rétro-propagation? Est-ce que la rétropropagation sert à appliquer la descente de gradient? Ou l'invers? En quoi l'un sert à l'autre, précisément?
16. Quelle est la différence entre descente de gradient Full/Stochastic/Batch? Cette différence est-elle seulement dans la procédure d'entraînement, ou y a-t-il aussi des différences dans la procédure de prédiction?
17. Définissez une époque d'entraînement.
18. Est-il important de mettre à l'échelle les données quand on utilise un réseau de neurones? Seulement quand on régularise le réseau de neurones ou même si on ne régularise pas?