

1. Quels sont les éléments essentiels qui ont accéléré les avancées en deep learning ?

- A) Des données variées, des capacités de calcul élevées, des algorithmes améliorés.
- B) La réduction du nombre de paramètres dans les modèles.
- C) L'augmentation de la taille des mini-lots uniquement.
- D) Uniquement l'utilisation des réseaux convolutifs.

2. Quelle fonction d'activation est souvent utilisée pour la classification multi-classes dans un réseau de neurones ?

- A) Sigmoidé
- B) Tanh
- C) ReLU
- D) Softmax

3. Dans la descente de gradient stochastique, les paramètres sont mis à jour :

- A) Après chaque lot de données
- B) Après chaque exemple d'entraînement
- C) Après chaque époque complète
- D) Jamais, ce sont les gradients qui changent

4. Quelle est l'utilité principale de la normalisation de mini-batch ?

- A) Accélérer l'apprentissage et réduire le surapprentissage
- B) Augmenter la complexité du modèle
- C) Simplifier la sélection des hyperparamètres
- D) Améliorer uniquement la précision finale du modèle

5. Quelle est la différence entre un paramètre et un hyperparamètre dans un réseau de neurones ?

- A) Les paramètres sont fixés par l'utilisateur, tandis que les hyperparamètres sont appris automatiquement.
- B) Les hyperparamètres sont ajustés par rétropropagation, contrairement aux paramètres.

- C) Les paramètres sont appris pendant l'entraînement, tandis que les hyperparamètres sont définis avant l'entraînement.
- D) Il n'y a pas de différence, ce sont des termes interchangeables.

6. La technique de "Dropout" en régularisation consiste à :

- A) Réduire la taille des mini-lots
- B) Ajouter des neurones au réseau pour plus de précision
- C) Ignorer aléatoirement certaines unités pendant l'entraînement pour éviter la co-dépendance
- D) Réduire les valeurs de poids pour éviter la saturation

7. Quelle est l'avantage de la descente de gradient par mini-lots ?

- A) Convergence plus rapide mais avec plus de bruit
- B) Compromis entre la stabilité et la rapidité de convergence
- C) Plus de stabilité avec un coût de calcul réduit
- D) Garantit une convergence plus précise

8. Pourquoi le taux d'apprentissage est-il réduit dynamiquement pendant l'entraînement ?

- A) Pour avancer plus lentement au début et plus rapidement à la fin
- B) Pour assurer une convergence stable et éviter les sauts autour du minimum global
- C) Pour augmenter la précision du modèle à chaque époque
- D) Pour réduire la complexité du modèle

9. Lorsqu'un modèle présente une variance élevée, quelle action est recommandée ?

- A) Augmenter la complexité du modèle
- B) Réduire la taille des mini-lots
- C) Ajouter plus de régularisation ou collecter plus de données d'entraînement
- D) Utiliser un autre algorithme d'optimisation

10. Dans une stratégie d'analyse d'erreurs, la première étape consiste à :

- A) Calculer la précision finale du modèle
- B) Collecter des exemples mal classifiés
- C) Ajuster les hyperparamètres
- D) Augmenter la taille de l'ensemble d'entraînement

11. Quel est l'objectif principal de la régularisation dans un réseau de neurones ?

- A) Augmenter la complexité du modèle
- B) Réduire le surapprentissage en limitant la capacité du modèle
- C) Accélérer l'entraînement en réduisant le nombre d'époques
- D) Ajuster automatiquement le taux d'apprentissage

12. Quelle est la particularité d'un réseau de neurones à faible profondeur ?

- A) Il est constitué de nombreuses couches cachées
- B) Il utilise la fonction Softmax dans toutes ses couches
- C) Il a une seule couche cachée et est adapté aux tâches simples
- D) Il est plus lent à entraîner mais offre une meilleure précision

13. Lorsqu'on utilise une descente de gradient par mini-lots, quel est l'impact d'une taille de mini-lot trop petite ?

- A) Elle peut entraîner des fluctuations de gradient trop importantes
- B) Elle stabilise l'apprentissage et améliore la convergence
- C) Elle permet d'éviter totalement le surapprentissage
- D) Elle augmente automatiquement le taux d'apprentissage

14. Dans le contexte de la configuration des jeux de données, que signifie un "biais élevé" ?

- A) Le modèle généralise bien mais sous-performe légèrement sur le test
- B) Le modèle n'apprend pas bien même sur les données d'entraînement, indiquant un sous-apprentissage
- C) Le modèle a des performances excellentes sur tous les jeux de données
- D) Le modèle présente un taux d'erreur élevé sur le jeu de test uniquement

15. Lors d'un ajustement manuel des hyperparamètres, quelle est la première étape recommandée ?

- A) Optimiser le nombre d'époques
- B) Fixer les hyperparamètres secondaires à des valeurs par défaut raisonnables
- C) Tester toutes les combinaisons d'hyperparamètres simultanément
- D) Diminuer progressivement le taux d'apprentissage

16. Parmi les algorithmes d'optimisation suivants, lequel est le plus adapté pour de grands ensembles de données et des projets complexes ?

- A) Descente de gradient par lots
- B) RMSprop
- C) Adam
- D) SGD avec Momentum

17. Pourquoi utiliser une fonction d'activation telle que ReLU dans les couches cachées d'un réseau de neurones ?

- A) Elle est particulièrement adaptée aux problèmes de classification multi-classes
- B) Elle permet d'introduire la non-linéarité, aidant le réseau à apprendre des relations complexes
- C) Elle normalise les activations dans chaque couche
- D) Elle ajuste automatiquement le taux d'apprentissage

18. Dans quel cas est-il recommandé d'utiliser la normalisation par mini-batch ?

- A) Lorsque l'on souhaite augmenter la variance dans les activations
- B) Pour stabiliser l'apprentissage en normalisant les activations de chaque mini-lot
- C) Pour réduire le nombre de paramètres dans le modèle
- D) Lorsqu'on utilise une descente de gradient par lots

19. Quelle est l'utilité de la technique de transfert d'apprentissage dans les projets de deep learning ?

- A) Elle permet de repartir l'entraînement à zéro avec des paramètres aléatoires
- B) Elle réduit le besoin de données en utilisant un modèle pré-entraîné
- C) Elle optimise automatiquement le taux d'apprentissage
- D) Elle permet d'entraîner le modèle uniquement sur des images

20. Quelle technique de descente de gradient permet généralement de trouver un compromis entre la stabilité de convergence et la rapidité d'entraînement ?

- A) Descente de gradient par lots
- B) Descente de gradient stochastique
- C) Descente de gradient par mini-lots
- D) Descente de gradient avec validation croisée