

Fiche d'auto-évaluation 05

May 30, 2020

Caractéristique = Feature

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

15. A-t-on la garantie qu'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.
16. Quelle est la relation entre descente de gradient et rétro-propagation?
17. Quelle est la différence entre descente de gradient Full/Stochastic/Batch? Cette différence se trouve seulement dans la procédure d'entraînement, ou il y a aussi des différences dans la procédure de prédiction?
18. Définissez une époque d'entraînement.
19. 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?
20. Jugez si la façon suivante de mettre à l'échelle (Min-Max scaling dans ce cas) est correct ou pas? Si elle n'est pas correcte, corrigez-la.
 - (a) On divise l'ensemble de données \mathcal{D} en données d'entraînement $\mathcal{D}^{\text{train}}$ et de test $\mathcal{D}^{\text{test}}$.
 - (b) On met les données de test à côté.
 - (c) On calcule la valeur minimale \min_j^{train} , maximale \max_j^{train} et moyenne μ_j^{train} de chaque colonne j sur les données d'entraînement.
 - (d) On transforme chaque valeur originale $x_j^{(i)}$ en $\frac{x_j^{(i)} - \dots}{\dots - \dots}$ (Completez la formule).
 - (e) On entraîne notre modèle sur l'ensemble $\mathcal{D}^{\text{train}}$ transformé.
 - (f) On prend les données de test $\mathcal{D}^{\text{test}}$, on calcule la valeur minimale \min_j^{test} , maximale \max_j^{test} et moyenne μ_j^{test} de chaque colonne j sur les données de test $\mathcal{D}^{\text{test}}$.
 - (g) On transforme $\mathcal{D}^{\text{test}}$ avec une formule similaire à avant, en utilisant cette fois \min_j^{test} , \max_j^{test} , μ_j^{test} .
 - (h) On utilise le modèle entraîné pour faire les prédictions sur $\mathcal{D}^{\text{test}}$ transformé.
21. Si on fait du 7-fold cross-validation, combien de modèles il faut entraîner?
22. Parlons de Sélection de Modèle (Model Selection): si on a un modèle et on veut trouver la meilleure combinaison de hyper-paramètres, comment peut-on utiliser la recherche par quadrillage (grid search) et cross-validation ensemble? Expliquez la procédure, pas à pas. (c'est la procédure implémentée par `sklearn.model_selection.GridSearchCV`)
23. Dans une forêt aléatoire, y-a-t-il des cas où augmenter le nombre d'arbre empire la justesse de la forêt? Si oui, quels cas?

24. Quelle est la différence entre une forêt aléatoire, bagging trees et extra-trees?
25. Si on a un ensemble d'arbres décisionnels, est-il préférable qu'ils se ressemblent ou qu'il soient différents?
26. Explique l'algorithme CART pour entraîner un arbre décisionnel, pas à pas. Après, explique comment on intègre cela dans une forêt aléatoire, en écrivant un pseudo-code.
27. Si on augmente le nombre d'arbres dans une forêt aléatoire, la variance de notre modèle diminue ou augmente?
28. Un arbre décisionnel, est-il un classificateur linéaire ou non? Et une forêt aléatoire?
29. Dans quel sens une prédiction d'une forêt aléatoire est interprétable? (Répond avec au moins deux éléments)
30. Comment peut-on calculer l'importance d'une colonne en se basant sur un ensemble d'arbres?
31. Est il nécessaire ou utile de mettre notre ensemble de données à l'échelle avant d'utiliser un arbre décisionnel ou un ensemble d'arbres?