

# Fiche d'auto-évaluation 03

January 18, 2022

1. Que veut dire qu'un modèle a une trop grande variance ? Et qu'est-ce que c'est la variance d'une caractéristique (écrivez la formule) ? La variance d'un modèle et la variance d'une caractéristique, sont-elles la même chose ? Où sont-elles des choses complètement différentes ?
2. Quelles techniques de sélection de caractéristiques (feature selection) nous avons vu jusqu'à présent ? Dis-en au moins 3.
3. Si on considère les types de modèles vus jusqu'à présent, quels sont les méthodes et les hyper-paramètres qui permettent d'augmenter ou diminuer leur complexité ?
4. Comment peut-on se rendre compte si un modèle est en surapprentissage (overfitting), ou en sous-apprentissage (underfitting) ? Comment peut-on le fixer dans les deux cas ?
5. À quel sert la régularisation ?
6. Montrez une procédure correcte pour choisir le bon coefficient de régularisation.
7. Dans le contexte d'une régression d'arête (Ridge regression), quels coefficients de régression obtenez-vous si vous fixez le coefficient de régularisation  $\alpha = 0$  ? Et si  $\alpha \rightarrow \infty$  ?
8. Pour régulariser un modèle, on va modifier la fonction de perte en ajoutant un "terme de régularisation". Est-ce cette affirmation correcte ? Si oui, faut-il ajouter ce terme au moment de l'entraînement, du test ou les deux ?
9. Montrez une procédure correcte pour choisir le bon coefficient de régularisation et le degré de la régression polynomiale.
10. Considérez un modèle de régression polynomiale et imaginez de augmenter le degré  $p$ . Est-ce que l'erreur sur les données d'entraînement augmente ou diminue ou on ne peut pas savoir ? Est-ce que l'erreur sur les données de test augmente ou diminue ou on ne peut pas savoir ?

11. Est la mise à l'échelle (scaling) importante/utile quand on fait de la régression linéaire? Pourquoi? Et quand on fait de la régression de Ridge (Ridge Regression, c-à-d régression + régularisation)? Pourquoi?
12. Considérez un modèle de régression linéaire et supposez que nos données ne sont pas passées en échelle. Peut-on affirmer que les caractéristiques avec les coefficients les plus petits sont les moins importantes? Imaginez maintenant de passer à l'échelle. Est-ce que l'affirmation précédente est correcte maintenant? Pourquoi? Faites un exemple.
13. Par rapport à la descente de gradient dans la régression logistique, le gradient dont on parle est le gradient de quelle fonction? Définissez cette fonction exactement.
14. La descente de gradient utilisée dans la régression logistique est utilisée pendant l'entraînement ou la prédiction?
15. Par rapport à la régression logistique, est-ce qu'elle donne directement la classe prédite ou une probabilité? Dans ce dernier cas, comment passe-t-on de la probabilité à la classe prédite?
16. Quelle est la différence entre la descente de gradient full/stochastique/batch?
17. Est la régression logistique un classificateur linéaire ou non? Que veut cela dire?
18. Que fait la fonction de softmax? Pourquoi est-elle employée? Écrivez sa formule.
19. Écrivez la formule de la cross-entropie.
20. Imaginez de travailler avec un ensemble de données caractérisé par un fort déséquilibre des classes et que vous entraînez directement un classificateur, sans tenir compte de ce déséquilibre. Imaginez d'obtenir une bonne justesse (accuracy). Feriez-vous confiance à ce classificateur? Pourquoi? Donnez un exemple où, même si la justesse du modèle est très haute, le modèle est mauvais.