

03. Self-assessment questions

May 3, 2021

- 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 ?
- Quelles techniques de sélection de caractéristiques (feature selection) nous avons vu jusqu'à présent ? Dis-en au moins 3.
- Dans les types de modèles vus jusqu'à présent, quels sont les méthodes et les hyper-paramètres qui permettent d'augmenter ou diminuer leur complexité ?
- 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 ?
- À quel sert la régularisation ?
- Montrez une procédure correcte pour choisir le bon coefficient de régularisation.
- Montrez une procédure correcte pour choisir le bon pair coefficient de régularisation et le degré de la régression polynomiale
- Is scaling strictly necessary or useful when applying linear regression? And when applying Ridge Regression? Why?
- For what concerns the gradient descent in logistic regression, we compute the gradient of which function, exactly?
- When using logistic regression, we use the gradient descent during training or during inference?
- In logistic regression, does the model give directly the predicted class of a sample or the probability? In the latter case, how do we go from the probability to the predicted class?
- Is logistic regression a linear or non-linear classifier? Why? What does "linear classifier" mean?

- What is the softmax function? Write its formula. Why is it employed?
- Write the formula of the cross-entropy.
- Suppose you have an unbalanced dataset and you directly train a model on top of it. Suppose you obtain a good accuracy, close to 99.9%. Would you say you have a good model? Why?