

03. Self-assessment questions

May 3, 2021

- What does it mean that a model has an excessive variance? What is, instead, the variance of a feature (write the formula)? Are the variance of a model and the variance of a feature the same thing? Or are they different concepts?
- Which methods to do feature selection have we seen up to now? Tell at least three of them.
- In the types of models you have seen so far, what are the methods and the hyper-parameters that allow to increase or decrease their complexity?
- How can you understand if a model is overfitting or underfitting? How can you improve the model, in case of overfitting or underfitting?
- What is the goal of regularization?
- Describe a correct procedure to choose a good regularization coefficient.
- In case of polynomial regression, describe a correct procedure to choose a good pair of regularization coefficient and polynomial degree
- Let us consider a polynomial regression model and suppose to increase the polynomial degree p . What happens to the training error? Does it increase or decrease or does it depend? And what about the test error?
- 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, the model gives 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?