

Revision questions for Chapter 4

Last updated: October 15, 2022

If you are asked to define some notion, you should explain carefully all notation (if any) that you use in your definition.

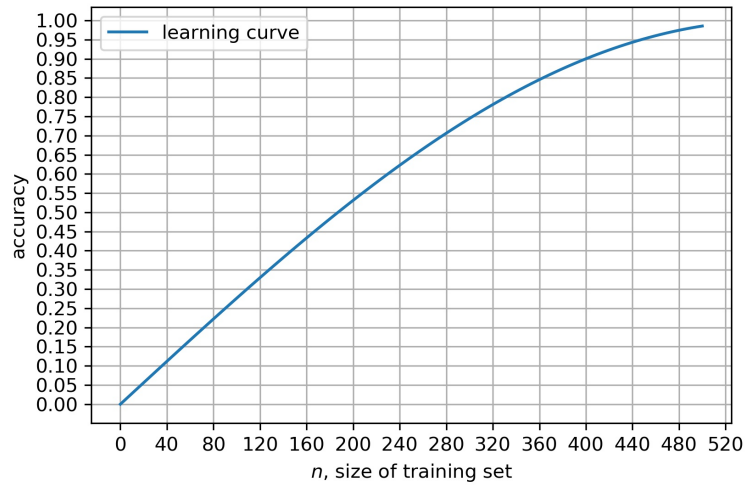
1. What is meant by *overfitting* in machine learning?
2. What is meant by *underfitting* in machine learning?
3. Define the *training accuracy* in machine learning.
4. Define the *generalization accuracy* in machine learning.
5. Compare and contrast the behaviour of the training and generalization accuracy as functions of model complexity.
6. How does the optimal model complexity depend on the size of the dataset?
7. What is meant by the *decision boundary* in machine learning?
8. Define *classifiers*, *regressors*, and *estimators* in the context of **scikit-learn**.
9. Give two strengths and two weaknesses of the K Nearest Neighbours algorithm.
10. Define the *learning curve* of a classifier.
11. Define the *learning curve* of a regressor.
12. Describe the method of *cross-validation* for evaluating generalization performance. Make sure to cover both the case of classification and the case of regression.
13. Explain what the following sequence of commands in **scikit-learn** is doing (assuming that all relevant modules have been imported):

```
iris = load_iris()
knn = KNeighborsClassifier(n_neighbors=3)
np.mean(cross_val_score(knn, iris.data, iris.target, cv=5))
```

Make sure to explain the role of options such as **cv=5**.

14. Explain why the cross-validation procedure leads to a downward bias in the estimate of the generalization performance.
15. Describe the method of *leave-one-out cross-validation* for evaluating generalization performance.
16. How does the bias of the K -fold cross-validation procedure depend on the number K of folds? Explain briefly why.

17. Estimate the downward bias of 10-fold cross-validation for a training set of size $n = 400$ and this learning curve:



18. Compare and contrast cross-validation and conformal prediction as answers to the question “How confident can we be in our prediction?”

Similar lists of questions will be produced for all chapters of the course to help students in revision. There is no guarantee that the actual exam questions will be in this list, or that they will be in any way similar.