Revision questions for Chapter 5

Last updated: January 28, 2022

If you are asked to define some notion, you should explain carefully all notation (if any) that you use in your definition. The question marked by (*) is more difficult. Answers to some questions are given in blue. All other answers can be found in the course notes (lecture slides or lab worksheets) provided on the course's Moodle page, in which case precise references are given (slide numbers are given according to the clean versions of the slides, those without solutions).

1. What is the linear regression model and what are its parameters? Chapter 5, slide 4.
2. Give an example of the parametric approach to machine learning and an example of the nonparametric approach to machine learning. Chapter 5, slide 5.
3. Give two advantages and a disadvantage of parametric methods. Chapter 5, slide 6.
4. When is the K Nearest Neighbours algorithm more likely to suffer from overfitting: for large or small values of K? When is the K Nearest Neighbours algorithm more likely to suffer from underfitting: for large or small values of K? Chapter 5, slide 7–8.
5. Define the Least Squares approach to linear regression. Chapter 5, slide 10.
6. What is meant by the residual sum of squares (RSS) in machine learning and statistics? Chapter 5, slide 10.
7. What is meant by *feature engineering* in machine learning? Chapter 5, slide 12.
8. You are given a classification problem with one feature x and the following training set:

x	y
-2	A
-1	A
0	B
2	B
4	C
5	C

As usual, y is the label. This is a multi-class classification problem with possible labels A, B, and C. The test samples are 0, 1, and -5.

Engineer an additional feature for this dataset, namely x^2. Therefore, your new training set still has 6 labelled samples in its training set and 3 unlabelled samples in its test set, but there are two features, x and x^2. Find the 1-Nearest Neighbour prediction for each of the test samples in the new dataset. Use the standard Euclidean metric. If you have encountered any ties, discuss briefly your tie-breaking strategy. The new dataset is

(x, x^2)	y
$(-2, 4)$	A
$(-1, 1)$	A
$(0, 0)$	B
$(2, 4)$	B
$(4, 16)$	C
$(5, 25)$	C

The test samples are $(0, 0)$, $(1, 1)$, and $(-5, 25)$. The (squared distances and) predictions are:

- For $(0, 0)$: prediction B (squared distances irrelevant).
- For $(1, 1)$: squared distances 18, 4, 2, 10, $9 + 15^2$, $4^2 + 24^2$; prediction B.
- For $(-5, 25)$: squared distances $3^2 + 21^2$, $4^2 + 24^2$, $5^2 + 25^2$, $7^2 + 21^2$, $9^2 + 9^2$, 10^2; prediction C.

Therefore, the predictions are: B for 0, B for 1, and C for -5. There were no ties.

9. Explain how a program for performing multiple linear regression (such as the function `LinearRegression` in `scikit-learn`) can be used for performing polynomial regression. Chapter 5, slide 13.

10. What is meant by the total sum of squares (TSS) in machine learning and statistics? Chapter 5, slide 14.

11. Define training R^2. What is its interpretation? Does $R^2 \approx 1$ mean good performance? Why? Can training R^2 be negative? Chapter 5, slide 15.

12. How is test R^2 defined in `scikit-learn`? Chapter 5, slide 17.

13. Can test R^2 be negative? If yes, explain why. If no, give an example. Chapter 5, slide 19.

14. Make sure you can do the exercise on slide 19 of Chapter 5.

15. Compare and contrast Least Squares and Ridge Regression as approaches to linear regression. Chapter 5, slides 21 and 23.

16. Explain the role of the parameter α for Ridge Regression. What happens when $\alpha = 0$? When $\alpha \to \infty$? Chapter 5, slide 24.

17. Compare and contrast Least Squares and the Lasso as approaches to linear regression. Chapter 5, slides 27 and 28.

18. Compare and contrast Ridge Regression and the Lasso as approaches to linear regression. Chapter 5, slides 27 and 28.

19. Why is the Lasso said to perform model selection? Because the Lasso sets many coefficients exactly to zero. Chapter 5, slides 27 and 30.

20. (*) Explain briefly why using the L_1 norm in the Lasso leads to many coefficients becoming exactly zero, unlike in the case of the L_2 norm used in Ridge Regression. Chapter 5, slides 31–33.

21. Briefly describe the method of elastic net in machine learning. Chapter 5, slide 34.

22. Briefly describe the kinds of datasets for which the Lasso would have an advantage over Ridge Regression, and the kinds of datasets for which Ridge Regression would have an advantage. Chapter 5, slide 37.

23. Give two advantages and two disadvantages of linear models. Chapter 5, slide 38.

24. Briefly describe the roles of the methods `__init__`, `fit`, and `predict` in the `Estimator` classes. What arguments do these methods typically require? Chapter 5, slides 41–42.

25. Give two examples of method chaining in `scikit-learn`. Chapter 5, slides 43–44.

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.