Revision questions for Chapter 7

Last updated: February 15, 2022

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

1. Define the *linear kernel*. Chapter 7, slide 4.

2. Define the *polynomial kernel* of degree d. Chapter 7, slide 5.

3. Define the *radial kernel*. What is its parameter? Chapter 7, slide 6.

4. Define the notion of a *kernel* in machine learning. Chapter 7, slide 8.

5. What is meant by the *kernel trick* in machine learning? Chapter 7, slide 9.

6. Give an advantage of using kernels over performing a feature mapping explicitly. Chapter 7, slide 10.

7. Consider the feature mapping given by

$$(x[0], x[1]) \mapsto (x[0]^2, x[1]^2, \sqrt{2}x[0]x[1]),$$

where $x[0]$ and $x[1]$ are real numbers (continuous features).

(a) What is the kernel function K for this feature mapping?
Answer: $K(x, x') = (x \cdot x')^2$ (or any other equivalent expression).
The calculation is

$$\begin{aligned}
&K((x[0], x[1]), (x'[0], x'[1])) \\
&= (x[0]^2, x[1]^2, \sqrt{2}x[0]x[1]) \cdot ((x'[0])^2, x'[1]^2, \sqrt{2}x'[0]x'[1]) \\
&= (x[0]x'[0])^2 + (x[1]x'[1])^2 + 2(x[0]x'[0]x[1]x'[1]) \\
&= (x[0]x'[0] + x[1]x'[1])^2 \\
&= ((x[0], x[1]) \cdot (x'[0], x'[1]))^2.
\end{aligned}$$

(b) What is $K((1, 2), (-3, 3))$?
Answer: 9.
The calculation is

$$((1, 2) \cdot (-3, 3))^2 = 3^2 = 9.$$

8. Give a necessary and sufficient condition for a continuous function K to be a kernel without using the notion of a feature mapping. Chapter 7, slide 10–12.

9. Prove that any kernel is symmetric and positive definite. Chapter 7, slide 13.

10. Let K_1 and K_2 be continuous kernels. Why is $K_1 + K_2$ a kernel as well? Chapter 7, slide 14. Use the answer to Question 8; it suffices to check that the sum of symmetric functions is symmetric and that the sum of positive definite functions is positive definite.

11. Let K be a continuous kernel and $w > 0$. Why is wK a kernel as well? Chapter 7, slide 14. Use the answer to Question 8; it suffices to check that the product of a scalar and a symmetric function is symmetric and that the product of a scalar and a positive definite function is positive definite.

12. Suppose K_1, K_2, and K_3 are continuous kernels. Why is their weighted average $0.1K_1 + 0.2K_2 + 0.7K_3$ always a kernel? Because adding and scaling continuous kernels give kernels. See Chapter 7, slide 15.

13. Suppose K is a kernel. What is the corresponding *normalized kernel*? What is the geometric meaning of normalization? Chapter 7, slide 16 (and also slide 17).

14. What is the normalized kernel corresponding to the polynomial kernel $K(x, x') = (1 + x \cdot x')^2$?

    Answer: it is

    $$\tilde{K}(x, x') = \frac{(1 + x \cdot x')^2}{\left(1 + \|x\|^2\right)\left(1 + \|x'\|^2\right)}.$$

15. What is the normalized kernel corresponding to the linear kernel $K(x, x') = x \cdot x'$? Make sure your definition works for all x and x', including zero vectors.

    For example, we can set

    $$\tilde{K}(x, x') = \begin{cases} \frac{x \cdot x'}{\|x\|\|x'\|} & \text{if } \|x\| > 0 \text{ and } \|x'\| > 0 \\ 0 & \text{otherwise.} \end{cases}$$

16. What is the role of the parameter of string kernels called the *decay factor*? Chapter 7, slide 19.

17. Define the feature mapping corresponding to string kernels. Chapter 7, slides 20–24.

18. Make sure you can do the exercises implicit in the examples on slides 20 and 23, including computing normalized kernels. Chapter 7, slides 20–23.

19. List the parameters of string kernels and briefly describe their role. Explain how the kernels corresponding to different values of the parameters can be combined. Chapter 7, slides 19 and 20. These are λ (the decay factor) and c (the length of the subsequences taken into account). Combination: Chapter 7, slide 24.

20. Give pseudocode for the kernelized version of the Nearest Neighbours algorithm. Chapter 7, slide 27.

21. Derive the kernel form of the Nearest Neighbours algorithm. Chapter 7, slide 27.

22. Consider the training set consisting of the following samples:

    - positive: $(1, -1, 0)$
    - positive: $(1, 2, -1)$
    - negative: $(-2, 1, 1)$
    - negative: $(-1, 1, 0)$.

    The test sample is $(0, 1, 0)$. Predict its label (assuming this is a classification problem) using the `K` Nearest Neighbours algorithm with the polynomial kernel $K(x, x') = (1 + x \cdot x')^2$, first for `K` $= 1$ and then for `K` $= 3$.

    Answer:

    - for `K` $= 1$: negative
    - for `K` $= 3$: positive

    The intermediate results are

$$
\begin{aligned}
&K(x^*, x^*) = 4 \\
&K(x_1, x_1) = 9 && K(x^*, x_1) = 0 \\
&K(x_2, x_2) = 49 && K(x^*, x_2) = 9 \\
&K(x_3, x_3) = 49 && K(x^*, x_3) = 4 \\
&K(x_4, x_4) = 9 && K(x^*, x_4) = 4 \\
&d^2(x^*, x_1) = 9 + 4 - 2 \times 0 = 13 \\
&d^2(x^*, x_2) = 49 + 4 - 2 \times 9 = 35 \\
&d^2(x^*, x_3) = 49 + 4 - 2 \times 4 = 45 \\
&d^2(x^*, x_4) = 9 + 4 - 2 \times 4 = 5,
\end{aligned}
$$

    where $x_1, \ldots, x_4$ are the training samples and x^* is the test sample.

These are the details of computations for one of the kernels above:

$K(x^*, x^*) = (1+x^* \cdot x^*)^2 = (1+(0,1,0)\cdot(0,1,0))^2 = (1+(0+1+0))^2 = 4.$

And the squared distances were computed using the formula

$$d^2(x, x') = K(x, x) + K(x', x') - 2K(x, x')$$

(see slide 27).

23. Make sure you can do the exercise on slide 29 of Chapter 7.

24. List three different practical applications of kernels. Chapter 7, slides 31–32.

25. Describe the `scikit-learn` class `KNeighborsClassifier` paying particular attention to the parameters `n_neighbors` and `metric` and methods `__init__`, `fit`, `predict`, `predict_proba`, and `score`. Lab Worksheet 7, Sections 1 and 4.

26. Describe the steps of creating your own estimator in `scikit-learn`. Lab Worksheet 7, Section 3.

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.