Revision questions for Chapter 7

Last updated: November 4, 2022

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

1. Define the *linear kernel*.
2. Define the *polynomial kernel* of degree d.
3. Define the *radial kernel*. What is its parameter?
4. Define the notion of a *kernel* in machine learning.
5. What is meant by the *kernel trick* in machine learning?
6. Give an advantage of using kernels over performing a feature mapping explicitly.
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?

   (b) What is $K((1,2),(-3,3))$?
8. Give a necessary and sufficient condition for a continuous function K to be a kernel without using the notion of a feature mapping.
9. Prove that any kernel is symmetric and positive definite.
10. Let K_1 and K_2 be continuous kernels. Why is $K_1 + K_2$ a kernel as well?
11. Let K be a continuous kernel and $w > 0$. Why is wK a kernel as well?
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?
13. Suppose K is a kernel. What is the corresponding *normalized kernel*? What is the geometric meaning of normalization?
14. What is the normalized kernel corresponding to the polynomial kernel $K(x, x') = (1 + x \cdot x')^2$?
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.
16. What is the role of the parameter of string kernels called the *decay factor*?

17. Define the feature mapping corresponding to string kernels.

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

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.

20. Give pseudocode for the kernelized version of the Nearest Neighbours algorithm.

21. Derive the kernel form of the Nearest Neighbours algorithm.

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 $\texttt{K} = 1$ and then for $\texttt{K} = 3$.

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

24. List three different practical applications of kernels.

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`.

26. Describe the steps of creating your own estimator in `scikit-learn`.

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.