Revision questions for Chapter 8

Last updated: November 10, 2022

The questions marked by (*) are more difficult. 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 the model for neural networks? You may assume that the number of hidden layers is 2. Describe its parameters.
2. Give an advantage and three disadvantages of neural networks.
3. What is a linear scoring function? How can it be used for classifying test samples into positive and negative?
4. Suppose we have a linear scoring function with parameters $b = -1$ and $w = (-2, 1, 0, 3)$. The test sample is $x^* = (0, 2, -1, 1)$. Calculate the predicted label for x^*.
5. How would you interpret the magnitude of a linear scoring function?
6. What is the *margin* of a given separating hyperplane?
7. Define the *maximum margin classifier*.
8. What is meant by the *optimal separating hyperplane*?
9. How is the optimal separating hyperplane used for classification?
10. Define the notion of a *support vector* in the context of maximum margin classifiers.
11. State an optimization problem whose solution is the maximum margin hyperplane. Give the geometric interpretation of this optimization problem.
12. Give an example of a training set for the problem of binary classification where no separating hyperplane exists.
13. Consider the training set consisting of three labelled samples, $(2, -1)$, $(3, -1)$, and $(0, 1)$. What is the optimal separating hyperplane? What is its margin?
14. In a binary classification problem, the training set consists of two labelled samples: positive $(1, 1)$ and negative $(0, 0)$. What is the optimal separating hyperplane? What is its margin?
    - Give an example of a non-optimal separating hyperplane. What is the margin of your hyperplane?
    - Give an example of a hyperplane that is not a separating hyperplane.

- Add one more labelled sample to the training set in such a way that it ceases to be linearly separable.

15. Consider the training set consisting of three labelled samples, $((1,4),-1)$, $((4,1),1)$, and $((1,-2),1)$.
    (a) Check that the hyperplane with $w=(0,-1)$ and $b=2$ separates this training set. Calculate its margin.
    (b) Draw the training samples and the hyperplane in $\mathbb{R}^2$.
    (c) From the drawing work out the optimal separating hyperplane and calculate its margin.
16. Give an example in which a separating hyperplane exists but a classifier based on a separating hyperplane is not desirable.
17. State the *soft margin classifier* as an optimization problem. Give the geometric interpretation of this problem.
18. What is (are) the parameter(s) of the soft margin classifier?
19. What is meant by the *slack variables* for the soft margin classifier?
20. How would you describe the role of the tuning parameter C in the soft margin classifier?
21. How is the maximum margin classifier a special case of the soft margin classifier?
22. Define the notion of a *support vector* in the context of a soft margin classifier.
23. (*) Consider the following training set:
    - positive samples: $(0,0)$, $(2,0)$, $(0,2)$, $(-2,0)$;
    - negative samples: $(3,0)$, $(0,-3)$, $(-3,2)$, $(0,4)$.

    Follow these steps to determine an optimal nonlinear separating curve for this training set:
    (a) Draw a graph showing all these samples. Can they be separated by a straight line?
    (b) What nonlinear transformation will allow us to represent a separating ellipse centred at the origin?
    (c) Transform the data from the original sample space to the new feature space using this transformation.
    (d) Draw a graph of the samples in the new feature space.
    (e) Determine the optimal separating hyperplane by inspection.
    (f) What is the equation of the optimal separating ellipse?

(g) Draw this ellipse on the graph you drew in step (a). Does the ellipse separate the points?

24. (*) Consider the same training set as in the previous question:

    - positive samples: $(0, 0)$, $(2, 0)$, $(0, 2)$, $(-2, 0)$;
    - negative samples: $(3, 0)$, $(0, -3)$, $(-3, 2)$, $(0, 4)$.

    Follow these steps to determine an optimal nonlinear separating line (not a straight line, of course) for this training set:

    (a) What nonlinear transformation will allow us to separate the positive samples and the negative samples? (Now there are simpler possibilities.)

    (b) Transform the data from the original sample space to the new feature space using this transformation.

    (c) Draw a graph of the samples in the new feature space.

    (d) Determine the optimal separating hyperplane by inspection.

    (e) What is the equation of the optimal separating line in the original space?

    (f) Draw this line on the graph you drew in step (a) of the previous question.

25. What is the main step (or the main steps) in the transition from the soft margin classifier to the support vector machine?

26. What is meant by a *support vector* in the context of support vector machines?

27. What is the role of kernels in support vector machines?

28. Give two examples of practical fields in which support vector machines are used.

29. Give an advantage and two disadvantages of support vector machines.

30. How would you use the support vector machine for classification problems with more than two classes?

    (a) Describe the one-vs-one approach.

    (b) Describe the one-vs-rest approach.

31. Make sure you can solve the exercise on slide 63 of Chapter 8.

32. Explain how the one-vs-rest procedure for multiclass classification can be used as inductive conformity measure.

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.