Revision questions for Chapter 8

Last updated: February 15, 2022

The question marked by (*) is more difficult. 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 (to the questions without (*)) can be found in the course notes (lecture slides or lab worksheets) provided on the course's Moodle page. The sign "=" is used for both precise and approximate equalities (feel free to do so when answering exam questions).

1. What is the model for neural networks? You may assume that the number of hidden layers is 2. Describe its parameters. Chapter 8, slides 4–9.

2. Give an advantage and three disadvantages of neural networks. Chapter 8, slides 10–11.

3. What is a linear scoring function? How can it be used for classifying test samples into positive and negative? Chapter 8, slide 22.

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^*.

   Answer: 1.

   The calculation is:

   $$w \cdot x^* = (-2) \times 0 + 1 \times 2 + 0 \times (-1) + 3 \times 1 - 1 = 4 > 0.$$

5. How would you interpret the magnitude of a linear scoring function? Chapter 8, slide 24.

6. What is the *margin* of a given separating hyperplane? Chapter 8, slide 26.

7. Define the *maximum margin classifier*. Chapter 8, slides 26–27.

8. What is meant by the *optimal separating hyperplane*? Chapter 8, slide 25.

9. How is the optimal separating hyperplane used for classification? Chapter 8, slide 27.

10. Define the notion of a *support vector* in the context of maximum margin classifiers. Chapter 8, slides 29–30.

11. State an optimization problem whose solution is the maximum margin hyperplane. Give the geometric interpretation of this optimization problem. Chapter 8, slides 31–32.

12. Give an example of a training set for the problem of binary classification where no separating hyperplane exists. Chapter 8, slide 35.

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? The optimal separating hyperplane is point 1. The margin is 1.

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? The optimal separating hyperplane is $x[0]+x[1]=1$. The margin is $1/\sqrt{2}=0.7071$.

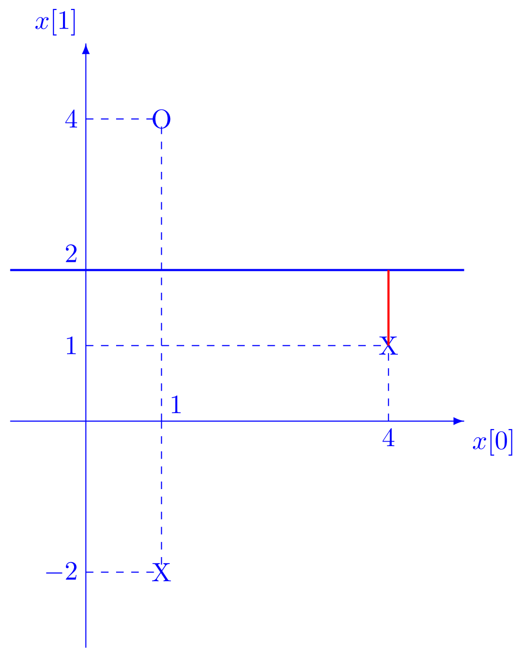
    - Give an example of a non-optimal separating hyperplane. What is the margin of your hyperplane?
    - Give an example of a hyperplane that is not a separating hyperplane.
    - Add one more labelled sample to the training set in such a way that it ceases to be linearly separable.

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.

    Answer: The margin is 1.

    It might be easiest to start from part (b) and read the margin off your drawing. My drawing is:



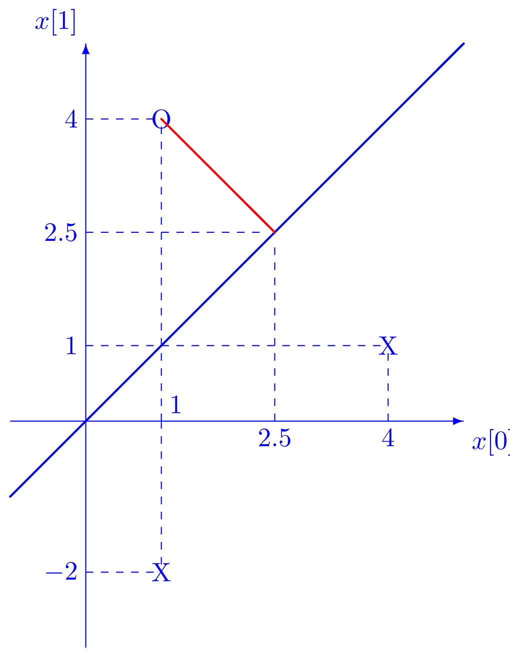
The equation of the hyperplane is $w[0]x[0] + w[1]x[1] + b = 0$, i.e., $2 - x[1] = 0$, i.e., $x[1] = 2$; it is shown as a thick blue line. The margin is represented by the thick red line. The separation is obvious.

(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.

Answer: the optimal separating hyperplane is $x[1] = x[0]$; the margin is $3/\sqrt{2} = 2.1213$.

My drawing is:



The thick blue line is the optimal separating hyperplane. The margin is represented by the thick red line (this choice is arbitrary to some degree; I could have chosen the symmetric line between the optimal separating hyperplane and X). The length of the red line is

$$\sqrt{1.5^2 + 1.5^2} = 3/\sqrt{2}.$$

16. Give an example in which a separating hyperplane exists but a classifier based on a separating hyperplane is not desirable. Chapter 8, slides 41–42.

17. State the *soft margin classifier* as an optimization problem. Give the geometric interpretation of this problem. Chapter 8, slides 46–48.

18. What is (are) the parameter(s) of the soft margin classifier? Chapter 8, slide 48.

19. What is meant by the *slack variables* for the soft margin classifier? Chapter 8, slide 46–47.

20. How would you describe the role of the tuning parameter C in the soft margin classifier? Chapter 8, slide 48.

21. How is the maximum margin classifier a special case of the soft margin classifier? The maximum margin classifier is the soft margin classifier for $C = \infty$, where C is the tuning parameter. Chapter 8, slide 48.

22. Define the notion of a *support vector* in the context of a soft margin classifier. Chapter 8, slide 49.

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? **Answer:** One possibility is to map $(x[0], x[1])$ to $(x[0]^2, x[1]^2)$.

    (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? **Answer**: $x[0]^2 + x[1]^2 = 6.5$ (if you chose the mapping given earlier).

    (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.) **Answer:** A simpler possibility is to map $(x[0], x[1])$ to $x[0]^2 + x[1]^2$. But probably the simplest possibility is to map $(x[0], x[1])$ to $|x[0]| + |x[1]|$.

    (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? **Answer**: $|x[0]|+|x[1]| = 2.5$ (if you chose the second mapping given earlier, $(x[0], x[1]) \mapsto |x[0]| + |x[1]|$).

(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? Chapter 8, slides 52–54.

26. What is meant by a *support vector* in the context of support vector machines? Chapter 8, slide 53.

27. What is the role of kernels in support vector machines? Chapter 8, slides 52–54.

28. Give two examples of practical fields in which support vector machines are used. Chapter 8, slide 57.

29. Give an advantage and two disadvantages of support vector machines. Chapter 8, slides 58–59.

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

    (a) Describe the one-vs-one approach. Chapter 8, slide 61.

    (b) Describe the one-vs-rest approach. Chapter 8, slide 64.

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. Chapter 8, slide 65.

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.