Revision questions for Chapter 3

Last updated: October 27, 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, and in many cases precise references are given (slide numbers are given according to the clean versions of the slides, those without solutions). The questions marked by (*) are more difficult; skip them unless you are aiming for a very high mark. The sign "=" is used for both precise and approximate equalities (feel free to do so when answering exam questions).

1. Give the definition of a conformity measure in the context of conformal prediction. Chapter 3, slide 8.

2. Give the definition of a nonconformity measure in the context of conformal prediction. Chapter 3, slides 8 and 28.

3. Define the notion of a conformal predictor. Chapter 3, slide 10.

4. Compare and contrast nonconformity measures and conformity measures. Chapter 3, slides 8 and 28.

5. In the context of conformal prediction, what is the minimal possible p-value for a training set of size n? Chapter 3, slide 11.

6. In the context of conformal prediction, what are the possible p-values for a training set of size n? Chapter 3, slide 11.

7. What is the main property of validity for conformal prediction? Chapter 3, slide 13.

8. What is meant by the efficiency of a conformal predictor? Chapter 3, slide 13.

9. Give three examples of conformity measures based on the method of Nearest Neighbours. Chapter 3, slide 15.

10. Consider the following training set in a multiclass classification problem:

    - samples of class A: $(-1, 0)$ and $(-1, -1)$;
    - samples of class B: $(0, 0)$ and $(0, 1)$;
    - samples of class C: $(1, 1)$, $(2, 1)$, and $(2, 0)$.

    The test sample is $(1, 0)$. When answering the following questions, use the conformity measure defined as the distance to the nearest sample of a different class divided by the distance to the nearest sample of the same class.

- What are the three p-values?
- What are the point prediction, confidence, and credibility?
- Suppose the label of the text sample is B. Compute the average false p-value.

This is similar to Chapter 3, slides 17–19.

First we assume that the label of $(1, 0)$ is A.

Sample	Label	Conformity score
$(-1, 0)$	A	$1/1 = 1$
$(-1, -1)$	A	$\sqrt{2}/1 = 1.414$
$(0, 0)$	B	$1/1 = 1$
$(0, 1)$	B	$1/1 = 1$
$(1, 1)$	C	$1/1 = 1$
$(2, 1)$	C	$\sqrt{2}/1 = 1.414$
$(2, 0)$	C	$1/1 = 1$
$(1, 0)$	A (?)	$1/2 = 0.5$

The test sample is the strangest (its rank is 1), and the p-value is $1/8 = 0.125$.

Next we assume that the label of $(1, 0)$ is B.

Sample	Label	Conformity score
$(-1, 0)$	A	$1/1 = 1$
$(-1, -1)$	A	$\sqrt{2}/1 = 1.414$
$(0, 0)$	B	$1/1 = 1$
$(0, 1)$	B	$1/1 = 1$
$(1, 1)$	C	$1/1 = 1$
$(2, 1)$	C	$\sqrt{2}/1 = 1.414$
$(2, 0)$	C	$1/1 = 1$
$(1, 0)$	B (?)	$1/1 = 1$

The rank of the test sample is 6, and the p-value is $6/8 = 0.75$.

Finally we assume that the label of $(1, 0)$ is C.

Sample	Label	Conformity score
$(-1, 0)$	A	$1/1 = 1$
$(-1, -1)$	A	$\sqrt{2}/1 = 1.414$
$(0, 0)$	B	$1/1 = 1$
$(0, 1)$	B	$1/1 = 1$
$(1, 1)$	C	$1/1 = 1$
$(2, 1)$	C	$2/1 = 2$
$(2, 0)$	C	$2/1 = 2$
$(1, 0)$	C (?)	$1/1 = 1$

The rank of the test sample is 5, and the p-value is $5/8 = 0.625$.

The point prediction is B, confidence is 0.375, and credibility is 0.75.

The average false p-value is 0.375 (computed as $(0.125 + 0.625)/2$).

11. How are conformity measures used for computing p-values in the context of conformal prediction? Chapter 3, slide 10.

12. How are nonconformity measures used for computing p-values in the context of conformal prediction? Chapter 3, slide 22.

13. Give at least two examples of nonconformity measures suitable for regression problems. Chapter 3, slide 25.

14. Discuss advantages and disadvantages of the conformity measures $\alpha_i = |y_i - \hat{y}_i|$ and $\alpha_i = |y_i - \hat{y}_i| / \sigma_i$, where $\hat{y}_i$ is a prediction for y_i and $\sigma_i > 0$ is an estimate of its accuracy. Chapter 3, slide 25.

15. How would you define a nonconformity measure based on the Nearest Neighbour algorithm and suitable for regression problems? Chapter 3, slide 26.

16. How would you define a nonconformity measure based on the K Nearest Neighbours algorithm and suitable for regression problems? Not on the slides (but see the previous question).

17. Write a pseudocode (or Python code) for using a grid for conformal prediction in the problem of regression. You may assume that the prediction set is an interval. Chapter 3, slide 31.

18. (*) Consider the training set

$$(x_1, y_1) = (0, 0), \qquad (x_2, y_2) = (2, 1), \qquad (x_3, y_3) = (1, 2),$$
$$(x_4, y_4) = (5, 7), \qquad (x_5, y_5) = (7, 7)$$

consisting of five training labelled samples in the problem of regression. Find the prediction set for the test sample $x^* = 6$ using the conformal predictor based on the 2 Nearest Neighbours algorithm with the nonconformity measure $\alpha_i = |y_i - \hat{y}_i|$, where $\hat{y}_i$ is the predicted label of x_i. As your significance level use $\epsilon = 20\%$. **Answer:** $[5.5, 8.5]$.

Solution: The nonconformity scores are:

Sample	Label	Prediction	NS
0	0	1.5	$\lvert 0 - 1.5\rvert = 1.5$
2	1	1	$\lvert 1 - 1\rvert = 0$
1	2	0.5	$\lvert 2 - 0.5\rvert = 1.5$
5	7	$(y + 7)/2$	$\lvert (y + 7)/2 - 7\rvert = \lvert (y - 7)/2\rvert$
7	7	$(y + 7)/2$	$\lvert (y + 7)/2 - 7\rvert = \lvert (y - 7)/2\rvert$
6	y	7	$\lvert y - 7\rvert$

The p-value is 20% or less if and only if the test labelled sample is the strangest (if it is the second strangest, the p-value is $2/6 > 20\%$). This is true for y satisfying

$$|y - 7| > |(y - 7)/2|$$

(which is equivalent to $y \neq 7$) and

$$|y - 7| > 1.5.$$

This gives the prediction interval $[5.5, 8.5]$.

19. Briefly explain how conformal prediction can be used for anomaly detection. Chapter 3, slide 35.

20. Define the point prediction, confidence, and credibility in the context of conformal prediction. Chapter 3, slide 39.

21. Let the size of the training set be n. Prove that the probability of error for a conformal predictor at significance level $\epsilon = 1/(n+1)$ does not exceed ϵ. (As usual in machine learning, you may make the IID assumption.) Chapter 3, slide 42.

22. Let the size of the training set be n. Prove that the probability of error for a conformal predictor at significance level $\epsilon = 2/(n+1)$ does not exceed ϵ. (This is the first exercise in Chapter 3, slide 43.) Follow the argument on slide 42:

    - What is the probability that $y^* \notin \Gamma^{2/(n+1)}$?
    - Notice that $y^* \notin \Gamma^{2/(n+1)}$ means that $z^* = (x^*, y^*)$ is the strangest or second strangest labelled sample in the set $z_1, \ldots, z_n, z^*$.
    - By the IID assumption, all permutations of $z_1, \ldots, z_n, z^*$ (and the corresponding permutations of $\alpha_1, \ldots, \alpha_n, \alpha_{n+1}$) have the same probability.
    - So the probability that the smallest or second smallest conformity score in the bag $\wr \alpha_1, \ldots, \alpha_{n+1} \int$ will be the last one is exactly $2/(n+1)$ (provided all αs are different; if the second and third smallest αs coincide, the probability of error will be less, either 0 or $1/(n+1)$).

23. When is a p-value regarded as statistically significant? highly statistically significant? Chapter 3, slide 44.

24. (*) Define randomized p-values in the context of conformal prediction. Chapter 3, slide 46.

25. (*) Define randomized prediction sets in the context of conformal prediction. Chapter 3, implicit on slide 46 (the same formula is used; see the bottom of slide 10).

26. (*) State the main property of validity of randomized conformal predictors in the online mode of prediction. Chapter 3, slides 47–48.

27. Define the average false p-value in the context of conformal prediction. Chapter 3, slides 49–50.

28. Make sure you can do the exercises given in the slides (slides 16–21, 24, 28–30, 33–34, 37–39, 50 of Chapter 3; some of these exercises are referred to as examples there).

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.