

Revision questions for Chapter 3

Last updated: April 25, 2022

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

1. Give the definition of a conformity measure in the context of conformal prediction.
2. Give the definition of a nonconformity measure in the context of conformal prediction.
3. Define the notion of a conformal predictor.
4. Compare and contrast nonconformity measures and conformity measures.
5. In the context of conformal prediction, what is the minimal possible p-value for a training set of size n ?
6. In the context of conformal prediction, what are the possible p-values for a training set of size n ?
7. What is the main property of validity for conformal prediction?
8. What is meant by the efficiency of a conformal predictor?
9. Give three examples of conformity measures based on the method of Nearest Neighbours.
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 test sample is B. Compute the average false p-value.
11. How are conformity measures used for computing p-values in the context of conformal prediction?

12. How are nonconformity measures used for computing p-values in the context of conformal prediction?
13. Give at least two examples of nonconformity measures suitable for regression problems.
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.
15. How would you define a nonconformity measure based on the Nearest Neighbour algorithm and suitable for regression problems?
16. How would you define a nonconformity measure based on the K Nearest Neighbours algorithm and suitable for regression problems?
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.
18. (*) Consider the training set

$$\begin{aligned} (x_1, y_1) &= (0, 0), & (x_2, y_2) &= (2, 1), & (x_3, y_3) &= (1, 2), \\ (x_4, y_4) &= (5, 7), & (x_5, y_5) &= (7, 7) \end{aligned}$$

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 $\epsilon = 20\%$.

19. Briefly explain how conformal prediction can be used for anomaly detection.
20. Define the point prediction, confidence, and credibility in the context of conformal prediction.
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.)
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 .
23. When is a p-value regarded as statistically significant? highly statistically significant?
24. (*) Define randomized p-values in the context of conformal prediction.
25. (*) Define randomized prediction sets in the context of conformal prediction.

26. (*) State the main property of validity of randomized conformal predictors in the online mode of prediction.
27. Define the average false p-value in the context of conformal prediction.
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.