Revision questions for Chapter 6

Last updated: October 28, 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.

1. What is meant by data normalization in machine learning? (Remember that in this course "normalization" is understood in the wide sense and includes the transformations perfomed by `Normalizer`, `StandardScaler`, etc., in `scikit-learn`.)

2. Why is normalization of features not essential for the method of Least Squares?

3. Why is normalization of features essential for Ridge Regression and the Lasso?

4. Briefly describe the class `StandardScaler` in `scikit-learn`, paying particular attention to its `fit` and `transform` methods.

5. Briefly describe the class `RobustScaler` in `scikit-learn`, paying particular attention to its `fit` and `transform` methods.

6. Briefly describe the class `MinMaxScaler` in `scikit-learn`, paying particular attention to its `fit` and `transform` methods.

7. Briefly describe the class `Normalizer` in `scikit-learn`, paying particular attention to its `fit` and `transform` methods.

8. Give an example of a dataset for which the use of the class `Normalizer` has a better justification than the use of classes perfoming normalization of features (such as `StandardScaler`).

9. Consider the following training set:

feature 1	feature 2	label
-3	2	male
0	5	female
3	8	male
0	8	male

   What is its normalized version, in the sense of `MinMaxScaler`? Apply the same transformation to the test set

feature 1	feature 2
1	-1
0	4
2	5

10. For the training set

feature 1	feature 2	label
-10	0	1.6
10	2	2.8

find its normalized version in the sense of `StandardScaler`. Apply the same transformation (emulating the `transform` method) to the test set

feature 1	feature 2
-20	-2
10	4
0	0

11. Consider the following training set:

feature 1	feature 2	label
-3	4	male
4	3	female
4	4	male

What is its normalized version, in the sense of `Normalizer`? Apply the same transformation to the test set

feature 1	feature 2
-4	3
3	-3

12. What is meant by data snooping in machine learning?

13. What is wrong with the following code for data normalization?

```
X = MinMaxScaler().fit_transform(boston.data)
X_train, X_test, y_train, y_test = train_test_split(X,
  boston.target)
```

Correct the code.

14. Discuss disadvantages of normalizing the training and test sets separately when using classes `StandardScaler`, `RobustScaler`, and `MinMaxScaler`.

15. Is it admissible to normalize training and test set separately when using the class `Normalizer`? Explain briefly why or why not.

16. What is wrong with the following code for data normalization?

```
X_train, X_test, y_train, y_test = train_test_split(X,
  boston.target)
X_train = MinMaxScaler().fit_transform(X_train)
X_test = MinMaxScaler().fit_transform(X_test)
```

Correct the code.

17. Explain the use of a validation set for parameter selection in machine learning.

18. What are disadvantages of the use of the test set for parameter selection (i.e., of choosing the parameters that give the best results on the test set)?

19. Explain the use of cross-validation for parameter selection in machine learning.

20. How would you perform parameter selection using grid search in a hierarchical manner to improve its computational efficiency?

21. Suppose the result of grid search using cross-validation to select parameters `C` and `gamma` is:

		`gamma`				
		0.5	1.0	1.5	2.0	2.5
	0.5	0.92	0.92	0.92	0.92	0.92
	1.0	0.92	0.92	0.92	0.92	0.92
`C`	1.5	0.92	0.92	0.92	0.92	0.92
	2.0	0.92	0.92	0.92	0.92	0.92
	2.5	0.92	0.92	0.92	0.92	0.92

(The entries in the table are the accuracy of the algorithm for different values of the parameters.) How suitable was this grid for selecting the optimal values of the two parameters? Explain briefly why.

22. Answer Question 21 for the following grid for parameters `A` and `B`:

		`B`				
		0.01	0.1	1	10	100
	0.01	0.59	0.66	0.74	0.73	0.77
	0.1	0.68	0.70	0.73	0.80	0.85
`A`	1	0.68	0.72	0.77	0.81	0.84
	10	0.69	0.77	0.80	0.85	0.89
	100	0.72	0.85	0.88	0.91	0.92

23. Answer Question 21 for the following grid for parameters `A` and `B`:

		`B`				
		0.01	0.1	1	10	100
	0.01	0.18	0.39	0.64	0.48	0.21
	0.1	0.36	0.58	0.80	0.61	0.43
`A`	1	0.63	0.82	0.95	0.83	0.57
	10	0.41	0.60	0.81	0.63	0.44
	100	0.23	0.44	0.63	0.47	0.31

24. List three desiderata for the method of inductive conformal prediction. Which of them are satisfied automatically?

25. Briefly explain why conformal prediction is not feasible in combination with feature normalization and parameter selection.

26. Compare and contrast conformal prediction and inductive conformal prediction.

27. Compare and contrast conformal prediction and cross-conformal prediction.

28. Compare and contrast inductive conformal prediction and cross-conformal prediction.

29. What is an *inductive conformity measure*? Define the inductive conformal predictor based on a given inductive conformity measure.

30. What is an *inductive nonconformity measure*? Define the inductive conformal predictor based on a given inductive nonconformity measure.

31. Give three examples of inductive nonconformity measures.

32. Give two examples of inductive conformity measures.

33. In the context of inductive conformal prediction, what is the minimal possible p-value for a training set proper of size $n - m$ and calibration set of size m?

34. Take the distance to the nearest neighbour (in the training set proper) of a different class as the conformity measure: $A(\zeta, (x, y))$ is the smallest distance from x to x' such that $(x', y') \in \zeta$ for some $y' \neq y$.

    - The training set proper is: -2 and -1 are labelled A, 2 is labelled B, and 6 is labelled C.
    - The calibration set is: -2 and 0 are labelled A, 3 is labelled B, and 5 and 7 are labelled C.
    - The test sample is 9.

    Compute the three p-values for the test sample. Summarize them in terms of the point prediction, confidence, and credibility.

35. In the case of regression, a simple inductive nonconformity measure is

    $$A(\zeta, (x, y)) = |y - \hat{y}|,$$

    where $\hat{y}$ is the prediction for the true label y computed from ζ as training set. Why is this inductive nonconformity measure sometimes regarded to be insufficiently flexible? How would you improve its flexibility?

36. Give a pseudo-code (or Python code) for the inductive conformal predictor based on the inductive conformity measure

$$A(\zeta, (x, y)) = |y - \hat{y}|$$

of the previous question.

37. (*) Prove that the code in Question 36 is correct.

38. Make sure you can do the exercise on slide 51 of Chapter 6.

39. Describe the `scikit-learn` class `GridSearchCV` paying particular attention to its methods `fit`, `predict`, and `score`.

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.