Revision questions for Chapter 9

Last updated: November 18, 2022

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

1. Consider the following code for estimating the prediction accuracy of the SVM on a training set `X_train`, `y_train` using feature normalization and parameter selection:

```
scaler = MinMaxScaler().fit(X_train)
X_train_scaled = scaler.transform(X_train)
X_test_scaled = scaler.transform(X test)
param_grid = {'C': [0.01, 0.1, 1, 10, 100],
    'gamma': [0.01, 0.1, 1, 10, 100]}
grid = GridSearchCV(SVC(), param_grid=param_grid)
grid.fit(X train_scaled, y_train)
print(grid.score(X_test_scaled, y_test))
```

   Locate data leakage in this code, and explain why it may lead to an overoptimistic estimate.

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

3. List two benefits of using the class `Pipeline`.

4. Describe the use of the `scikit-learn` class `Pipeline` for parameter selection using a grid search.

5. Define the cross-conformal predictor based on a given inductive conformity measure.

6. Explain why the requirement of validity may be violated for cross-conformal predictors.

7. Make sure you can solve the exercise implicit in the example on slide 29 of Chapter 9.

8. Define the *calibration curve* for a set predictor.

9. Describe the class `KFold` in `scikit-learn`.

10. Describe the method `decision_function` in the class `SVC` and explain how it can be useful in implementing inductive conformal predictors and cross-conformal predictors.

Similar lists of questions will be produced for all chapters of the module 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.