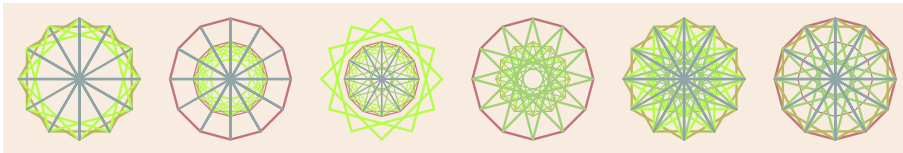


Review

CSC316 Machine Learning Professor Leon Tabak

16 February 2022

This work is licensed under CC BY 4.0. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.



1. Find an alternative to StandardScaler in the scikit-learn library.
2. For what purpose did we use an imputer?
3. Find a way to count the number of NaNs in a one dimensional numpy array.
4. What methods must a CustomTransformer have?
5. How are the Pipeline and ColumnTransformer classes related?
6. We replaced missing values in a column with the column's median value. What other choices did we have?
7. We split our dataset into a training set and a test set. Then we created validation sets. Did we take the validation sets from the training set or the test set?
8. What is the purpose of GridSearchCV?
9. Instead of displaying a confusion matrix as a square table of numbers, we can show a square array of tiles in which each tile is drawn with a shade of gray.
If our model is working perfectly, what will the picture of the confusion matrix look like?

10. The MNIST dataset has some features that we will find convenient. Explain.
11. We are looking at binary classifier that uses the SGDClassifier. Each letter in “SGD” is the first letter of a word. What are those three words? What do they mean?
12. What is another word for True Negative Rate (TNR)? (It begins with an “s.”)
13. What is another word for recall? (It begins with an “s.”)
14. Let’s suppose that we are building a program that will classify objects. Each object belongs to one of N classes. We are going to use a “OvO” (one versus one) strategy.

This means that we need a binary classifier for each possible pairings of distinct classes.

If we are trying to identify the birds that come to the bird feeder and the only kinds of birds that visit are sparrows, finches, and chickadees, then we need 3 binary classifiers. . .

- a sparrow versus finch classifier
- a sparrow versus chickadee classifier
- a finch versus chickadee classifier

If there are N classes, how many binary classifiers do we need?