

Feature Clustering

Use agglomerative clustering for feature dimensionality reduction.

Chapter Goals:

• Learn how to use agglomerative clustering for feature dimensionality reduction

A. Agglomerative feature clustering

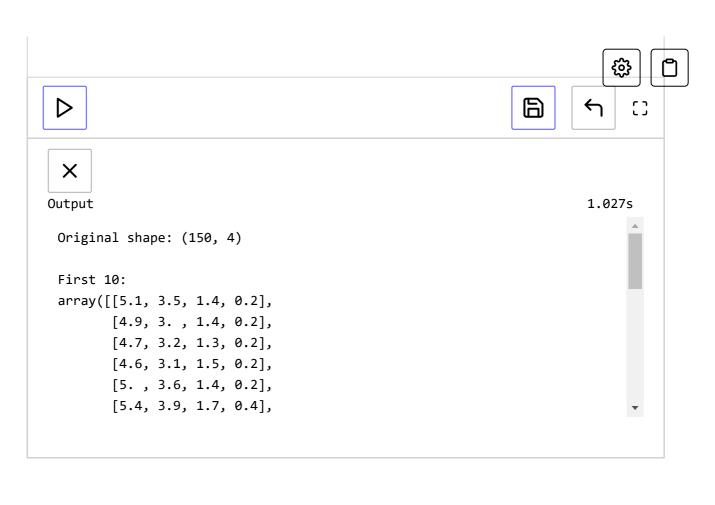
In the **Data Preprocessing** section, we used PCA to perform feature dimensionality reduction on datasets. We can also perform feature dimensionality reduction using agglomerative clustering. By merging common features into clusters, we reduce the number of total features while still maintaining most of the original information from the dataset.

In scikit-learn, we perform agglomerative clustering on features using the FeatureAgglomeration (https://scikit-

learn.org/stable/modules/generated/sklearn.cluster.FeatureAgglomeration .html#sklearn.cluster.FeatureAgglomeration) object (part of the cluster module). When initializing the object, n_clusters keyword argument (which represents the number of final clusters) is used to specify the new feature dimension of the data.

The code below demonstrates how to use the FeatureAgglomeration object to reduce feature dimensionality from 4 to 2. We use the object's fit_transform function to fit the clustering model on the data, then subsequently apply the feature reduction on the data.

```
1 # predefined data
2 print('Original shape: {}\n'.format(data.shape))
3 print('First 10:\n{}\n'.format(repr(data[:10])))
4
5 from sklearn.cluster import FeatureAgglomeration
6 agg = FeatureAgglomeration(n_clusters=2)
7 new_data = agg.fit_transform(data)
8 print('New shape: {}\n'.format(new_data.shape))
9 print('First 10:\n{}\n'.format(repr(new_data[:10])))
```





Evaluating Clusters

✓ Mark as Completed

Quiz

64% completed, meet the criteria and claim your course certificate!

Buy Certificate

Properties (1) Report an Issue

? Ask a Question

 $(https://discuss.educative.io/tag/feature-clustering__clustering-with-scikit-learn__machine-learning-for-software-engineers)\\$