The preprocessing technique in which a dataset is transformed to a distribution of mean 0 and variance 1 is known as _____.
Mean Removal

What is the output of the following code?
```
import sklearn.preprocessing as preprocessing
x = [[7.8], [1.3], [4.5], [0.9]]
print(preprocessing.Binarizer().fit(x).transform(x).shape)
```
(4, 1)

The preprocessing technique in which categorical values are transformed to categorical integers is known as _____.
Encoding

What is the output of the following code?
```
import sklearn.preprocessing as preprocessing
x = [[0, 0], [0, 1], [2,0]]
enc = preprocessing.OneHotEncoder()
print(enc.fit(x).transform([[1, 1]]).toarray())
```
[[0. 0. 0. 1.]]

Which of the following module of sklearn contains preprocessing utilities?
Preprocessing

Which of the following API is used to scale a dataset to range 0 and 1?
MinMaxScaler

What is the output of the following code?
```
import sklearn.preprocessing as preprocessing
regions = ['HYD', 'CHN', 'MUM', 'HYD', 'KOL', 'CHN']
print(preprocessing.LabelEncoder().fit(regions).transform(regions))
```
[1 0 3 1 2 0]

The preprocessing technique in which missing values are replaced with the mean of a dataset is known as _____.
Imputing

What is the output of the following code?
```
import sklearn.preprocessing as preprocessing
x = [[7.8], [1.3], [4.5], [0.9]]
print(preprocessing.Binarizer().fit(x).transform(x))
```
[[1.]
 [1.]
 [1.]
 [1.]]

Scikit-learn provides Pipeline utility to build a pipeline, which performs a series of transformations.
True

_____ parameter is used to control the number of neighbors of KNearestClassifier.
n_value

Which regressor utility of sklearn.neighbors is used to learn from k nearest neighbors of each query point?
KNeighborsRegressor

Which of the following module of sklearn is used to deal with Nearest Neighbors?
n_neighbors

Which of the following parameter can be used to give more weightage to the points, which are nearer to a point in the nearest neighbors method?
weights

This study source was downloaded by 100000850522783 from CourseHero.com on 07-02-2023 01:05:10 GMT -05:00

https://www.coursehero.com/file/52381959/Machine-Learning-Using-scikit-learntxt/

What is the strategy followed by Radius Neighbors method?
It looks in the vincinity of area, covered by a fixed radius, of each training point.

Which of the following class is used to implement the K-Nearest Neighbors classification in scikit-learn?
KNeighborsClassifier

Neighbors-based regression is mainly used when the data labels are continuous rather than discrete variables.
True

Which of the following algorithms can be used with any nearest neighbors utility in scikit-learn?
all

A feature can be reused to split a tree during Decision tree creation.
True

Which of the following module of sklearn is used for dealing with Decision Trees?
tree

Which of the following utility is used for regression using decision trees?
DecisionRegressor

Decision trees overfit the data very easily.
True

Which of the following parameter is used to tune a Decision Tree?
split

Data used for Decision Trees have to be preprocessed compulsorily.
False

A small change in data features may change a Decision Tree completely.
True

Ensemble methods are better than Decision Trees.
True

More improvement is found in an ensemble when base estimators are highly correlated?
False

Which of the following utility of sklearn.ensemble is used for classification with extra randomness?
RandomForestClassifier

Which of the following utility of sklearn.ensemble is used for implementing classification with the bagging method?
BaggingClassifier

Which of the following are Boosting ensemble methods?
Adaboost, Gradient Tree Boosting

Which of the following module of sklearn is used for dealing with ensemble methods?
ensemble

Which parameter is used to manage many base estimators in RandomForestClassifier?
n_estimators

This study source was downloaded by 100000850522783 from CourseHero.com on 07-02-2023 01:05:10 GMT -05:00

https://www.coursehero.com/file/52381959/Machine-Learning-Using-scikit-learntxt/

Which approach is used by SVC and NuSVC for multi-class classification?
one vs one

What happens when very small value is used for parameter C in support vector machines?
Misclassification happens- but in here all options are wrong

Which of the following module of sklearn provides the utilities to deal with support vector machines?
svm

SVM algorithms are memory efficient.
True

Which attribute provides details of obtained support vectors, after classifying data using SVC?
support_vectors
support_vectors_

What values can be used for kernel parameter of SVC class?
All

LinearSVC class accepts kernel parameter value.
False

Scaling or Normalization of data improves the accuracy of support vector machines.
True

Which of the following parameter of SVC method is used for fine-tuning the model?
C

Which of the following utilities are provided by sklearn to perform classification using support vector machines?
All

Agglomerative Clustering follows a top-down approach.
false

Which of the following attribute is used to access cluster centers, after completing clustering of given data points, using one of the clustering algorithms?
cluster_centers_

Which of the following parameters are used to control Density-based clustering?
eps, min_samples

Which of the following parameters are used to control Affinity Propagation clustering ?
preference, damping

Which of the following clustering technique is used to group data points into user given k clusters?
K-means clustering

What values can be used for the linkage parameter in AgglomerativeClustering?
All

Which of the following utility of sklearn.cluster is used for performing k-means clustering?
KMeans()

This study source was downloaded by 100000850522783 from CourseHero.com on 07-02-2023 01:05:10 GMT -05:00

https://www.coursehero.com/file/52381959/Machine-Learning-Using-scikit-learntxt/

Spectral Clustering is best suited for identifying dense clusters.
False

What does the Homogeneity score of a clsutering algorithm indicate ?
Verifies if each cluster contains only members of a single class.\

Which of the following module of sklearn contains popular processed datasets?
datasets

Data used for Decision Trees have to be preprocessed compulsorily.
False

Which of the following API is used to normalize a sample to the unit norm?
Normalizer

Which of the following library is widely used to read data from external sources with structured data?
pandas

Which of the following expressions can access the features of the iris dataset, shown in the below expression?
```
from sklearn import datasets
iris = datasets.load_iris()
iris.data
```

What do the methods starting with fetch, of sklearn.datasets module do?
Downloads a specific dataset from a library

Which of the following is an important parameter of RadiusNeighborsClassifier?
radius

This study source was downloaded by 100000850522783 from CourseHero.com on 07-02-2023 01:05:10 GMT -05:00

https://www.coursehero.com/file/52381959/Machine-Learning-Using-scikit-learntxt/