Experiment :3

Problem statement:

Implement k-nearest neighbours classification using python

Aim: to Implement k-nearest neighbours classification using python

ALGORITHM:

Step 1: Load the data

Step 2: Initialize the value of k

Step 3: For getting the predicted class, iterate from 1 to total number of training data points

i) Calculate the distance between test data and each row of training data. Here we will use Euclidean distance as our distance metric since it's the most popular method. The other metrics that can be used are Chebyshev, cosine, etc.

ii) Sort the calculated distances in ascending order based on distance
values 3. Get top k rows from the sorted array

iii) Get the most frequent class of these rows i.e. Get the labels of the selected K entries

iv) Return the predicted class If regression, return the mean of the K labels If classification, return the mode of the K labels

- If regression, return the mean of the K labels
- If classification, return the mode of the K labels

Step 4: End.

PROGRAM:

```
import numpy as np
from sklearn import datasets
iris = datasets.load_iris()
data = iris.data
labels = iris.target
for i in [0, 79, 99, 101]:
 print(f"index: {i:3}, features: {data[i]}, label: {labels[i]}")
np.random.seed(42)
indices = np.random.permutation(len(data))
n_training_samples = 12
learn_data = data[indices[:-n_training_samples]]
learn_labels = labels[indices[:-n_training_samples]]
test_data = data[indices[-n_training_samples:]]
test_labels = labels[indices[-n_training_samples:]]
print("The first samples of our learn set:")
print(f"{'index':7s}{'data':20s}{'label':3s}")
for i in range(5):
 print(f"{i:4d} {learn_data[i]} {learn_labels[i]:3}")
 print("The first samples of our test set:")
 print(f"{'index':7s}{'data':20s}{'label':3s}")
for i in range(5):
 print(f"{i:4d} {learn_data[i]} {learn_labels[i]:3}")
```

OUTPUT:

```
index:   0, features: [5.1 3.5 1.4 0.2], label: 0
index:  79, features: [5.7 2.6 3.5 1. ], label: 1
index:  99, features: [5.7 2.8 4.1 1.3], label: 1
index: 101, features: [5.8 2.7 5.1 1.9], label: 2

The first samples of our learn set:
index  data           label
  0 [6.1 2.8 4.7 1.2]   1
  1 [5.7 3.8 1.7 0.3]   0
  2 [7.7 2.6 6.9 2.3]   2
  3 [6.  2.9 4.5 1.5]   1
  4 [6.8 2.8 4.8 1.4]   1

The first samples of our test set:

index  data           label
  0 [6.1 2.8 4.7 1.2]   1
  1 [5.7 3.8 1.7 0.3]   0
  2 [7.7 2.6 6.9 2.3]   2
  3 [6.  2.9 4.5 1.5]   1
  4 [6.8 2.8 4.8 1.4]   1
```

Result: The program has been executed successfully and data classified using KNN algorithm.