Machine Learning
Experiment 5

SAP ID : 60004200139   Name : Riya Bihani

Division : B   Batch : B1

AIM

To implement K-Nearest Neighbour.

THEORY

K-Nearest Neighbours (KNN) is one of the most basic yet essential classification algorithms in Machine Learning. It is based on supervised learning technique. It assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. It stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.

K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. It is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. At the training phase, it just stores the dataset and when it gets new data, then it classifies that data into a category that is much like the new data.

Example - Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So, for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.

CODE

```
import math

interview = [70, 70, 30, 10]
exam_rank = [70, 40, 40, 40]
classes = ['not hired', 'hired', 'not hired', 'not hired']
data = {
   'first' : [70, 70, 'not hired'],
   'second' : [70, 40, 'hired'],
   'third' : [30, 40, 'not hired'],
```

```python
    'fourth' : [10, 40, 'not hired'],
}
# print(data)

x = 30
y = 70

distance1 = []
distance2 = []
response = []

for i in range(len(interview)) :
  s = (interview[i] - x)**2 + (exam_rank[i] - y)**2
  s = math.sqrt(s)
  # print(s)
  distance1.append(s)
  distance2.append(s)

# print('Euclidean Distance - ', distance1)
distance2.sort()
# print(distance2)

for i in range(len(distance1)):
  for j in range(len(distance2)):
    if(distance1[i] == distance2[j]):
      if j == 0:
        response.append(data['first'][2])
      elif j == 1:
        response.append(data['second'][2])
      elif j == 2:
        response.append(data['third'][2])
      elif j == 3:
        response.append(data['fourth'][2])

# print(response)

k = 3
countNH = countH = 0
for i in range(k):
  if response[i] == 'not hired':
    countNH += 1
```

```
  else:
    countH += 1

# print(countNH, countH)

if countNH > countH:
  print('Type for {} and {} will be not hired.'.format(x, y))
else:
  print('Type for {} and {} will be hired.'.format(x, y))
```

OUTPUT

```
Type for 30 and 70 will be not hired.
```

CONCLUSION

Hence, we have successfully implemented K-Nearest Neighbours in classification.