University of Canberra

Faculty of Science and Technology

**Programming for Data Science G (11521)**

**Answer to Week 9 Tutorial Question**

**Vector Quantisation and Nearest Centroid Classifier (or Nearest Prototype Classifier)**

**Your task: Implement the following program**

Repeat the following steps 3 times with K = 2, 3, and 4:

* Run **K-means Clustering** program to cluster the given blue and red datasets in Assignment 1 and output their cluster centres to file as follows
  + blue\_2d.txt 🡪 K-means clustering 🡪 blue\_model\_2d.txt
  + red\_2d.txt 🡪 K-means clustering 🡪 red\_model\_2d.txt
  + blue\_4d.txt 🡪 K-means clustering 🡪 blue\_model\_4d.txt
  + red\_4d.txt 🡪 K-means clustering 🡪 red\_model\_4d.txt
  + blue\_8d.txt 🡪 K-means clustering 🡪 blue\_model\_8d.txt
  + red\_8d.txt 🡪 K-means clustering 🡪 red\_model\_8d.txt
* Run **Nearest Neighbour Classifier** program on the given unknown datasets and the model datasets output from the K-Means Clustering program as follows
  + blue\_model\_2d.txt, red\_model\_2d.txt, and unknown\_2d.txt 🡪 Nearest Neighbour Classifier 🡪 results
  + blue\_model\_4d.txt, red\_model\_4d.txt, and unknown\_4d.txt 🡪 Nearest Neighbour Classifier 🡪 results
  + blue\_model\_8d.txt, red\_model\_8d.txt, and unknown\_8d.txt 🡪 Nearest Neighbour Classifier 🡪 results
* Compare the above results with the results you got from the red, blue, and unknown datasets for Assignment 1.

**Answer:**

import io\_data\_module as iodata

#Function to build model file

def build\_model(K, data\_filename, model\_filename):

cluster\_centre\_list = iodata.run\_kmeans(K, data\_filename)

#print(cluster\_centre\_list)

f = open(model\_filename, 'w')

for centre in cluster\_centre\_list:

output\_line = ''

for element in centre[:-1]:

output\_line += str(element) + ' '

output\_line += str(centre[-1]) + '\n'

f.write(output\_line)

f.close()

#end of function

#Main program

blue\_files = ['datasets/blue\_2d.txt', 'datasets/blue\_4d.txt', 'datasets/blue\_8d.txt']

blue\_model\_files = ['datasets/blue\_model\_2d.txt', 'datasets/blue\_model\_4d.txt', 'datasets/blue\_model\_8d.txt']

red\_files = ['datasets/red\_2d.txt', 'datasets/red\_4d.txt', 'datasets/red\_8d.txt']

red\_model\_files = ['datasets/red\_model\_2d.txt', 'datasets/red\_model\_4d.txt', 'datasets/red\_model\_8d.txt']

unknown\_files = ['datasets/unknown\_2d.txt', 'datasets/unknown\_4d.txt', 'datasets/unknown\_8d.txt']

for K in range(2, 5):

for index in range(len(blue\_files)):

print('K =', K)

print('Files: ', blue\_model\_files[index], red\_model\_files[index], unknown\_files[index])

build\_model(K, blue\_files[index], blue\_model\_files[index])

build\_model(K, red\_files[index], red\_model\_files[index])

result = iodata.run\_nearest\_neighbour(blue\_model\_files[index], red\_model\_files[index], unknown\_files[index])

print(result)

**Output:**

K = 2

Files: datasets/blue\_model\_2d.txt datasets/red\_model\_2d.txt datasets/unknown\_2d.txt

0.678713 0.951598 blue

-1.169512 -0.957855 blue

0.631947 1.236489 blue

-0.131799 -0.324218 blue

-0.199565 -0.229171 blue

0.900907 0.448131 blue

0.699395 0.414314 blue

-1.132646 -1.013298 blue

1.004178 1.355361 blue

-0.127378 -0.097121 blue

6.08804 3.457729 red

4.147974 5.275341 red

6.538759 3.670323 red

4.579573 4.03559 red

4.756026 4.184762 red

5.221742 2.872705 red

5.271773 3.158064 red

4.046376 5.19232 red

6.530952 3.171413 red

4.918007 4.142507 red

K = 2

Files: datasets/blue\_model\_4d.txt datasets/red\_model\_4d.txt datasets/unknown\_4d.txt

5.5 2.4 3.8 1.1 red

5.7 2.8 4.1 1.3 red

5.8 2.6 4.0 1.2 red

6.2 2.9 4.3 1.3 red

6.1 2.7 4.3 1.3 red

5.5 2.4 3.7 1.0 red

5.8 2.7 3.9 1.2 red

6.0 2.7 5.1 1.6 red

5.4 3.0 4.5 1.5 red

6.3 2.3 4.4 1.3 red

5.6 3.0 4.1 1.3 red

5.5 2.5 4.0 1.3 red

5.5 2.6 4.4 1.2 red

6.1 3.0 4.6 1.4 red

5.0 2.3 3.3 1.0 blue

5.6 2.7 4.2 1.3 red

5.7 3.0 4.2 1.2 red

5.7 2.9 4.2 1.3 red

5.1 2.5 3.0 1.1 blue

6.0 3.4 4.5 1.6 red

K = 2

Files: datasets/blue\_model\_8d.txt datasets/red\_model\_8d.txt datasets/unknown\_8d.txt

-1.252161 4.595982 0.403397 4.664438 1.184887 4.729431 0.483914 4.661279 blue

1.058662 4.139604 -0.743436 5.450736 -0.274353 4.145079 -0.467963 4.333563 blue

-0.835754 4.782662 -1.13831 5.45584 0.135823 4.216405 -0.231157 5.910545 blue

1.603099 4.092962 0.79424 5.514881 -1.259234 5.508506 -0.812622 4.289069 blue

-0.288341 4.550324 -0.597914 4.54299 -1.532651 5.006298 0.86868 4.734013 blue

-0.918091 5.803641 -0.406679 4.369684 -0.598748 5.700556 -0.294587 4.781506 blue

0.920304 4.52264 0.313416 4.17575 -1.23142 4.807567 0.873939 4.097733 blue

-1.892713 5.419602 0.145955 5.497594 -0.847049 4.472751 0.649144 5.086061 blue

0.990046 4.553275 0.740694 4.660983 -0.074056 4.290081 -0.216103 4.231506 blue

-0.623098 4.631027 0.470979 5.102245 1.139348 5.049803 1.562672 4.929757 blue

-1.917901 1.557697 1.61902 2.636847 0.345654 2.062426 0.32132 2.998103 red

-0.385049 3.148221 -0.72842 1.82968 -0.049924 2.331142 0.851898 2.402296 red

1.576337 2.411215 -0.884658 2.606353 -1.166365 2.880788 0.888455 1.649488 red

0.300287 1.865714 -0.168243 2.290775 1.491714 2.283442 -1.825572 2.360293 red

0.510346 1.917645 0.657498 1.74974 1.799944 2.373934 -1.568937 2.991486 red

-0.586001 2.052022 -0.665066 3.081557 0.885694 1.791796 1.468184 2.046193 red

1.818772 2.372645 -1.262745 2.019086 0.305039 2.753358 -0.487803 1.778518 red

1.358644 2.984491 -0.061581 1.98725 0.688312 1.813675 0.969963 2.196707 red

0.246576 2.359044 -1.153771 2.375196 0.141793 2.199259 0.52315 3.053531 red

0.374774 2.540291 -0.706522 2.680273 -1.404407 2.011466 1.878639 2.524433 red

K = 3

Files: datasets/blue\_model\_2d.txt datasets/red\_model\_2d.txt datasets/unknown\_2d.txt

0.678713 0.951598 blue

-1.169512 -0.957855 blue

0.631947 1.236489 blue

-0.131799 -0.324218 blue

-0.199565 -0.229171 blue

0.900907 0.448131 blue

0.699395 0.414314 blue

-1.132646 -1.013298 blue

1.004178 1.355361 blue

-0.127378 -0.097121 blue

6.08804 3.457729 red

4.147974 5.275341 red

6.538759 3.670323 red

4.579573 4.03559 red

4.756026 4.184762 red

5.221742 2.872705 red

5.271773 3.158064 red

4.046376 5.19232 red

6.530952 3.171413 red

4.918007 4.142507 red

K = 3

Files: datasets/blue\_model\_4d.txt datasets/red\_model\_4d.txt datasets/unknown\_4d.txt

5.5 2.4 3.8 1.1 red

5.7 2.8 4.1 1.3 red

5.8 2.6 4.0 1.2 red

6.2 2.9 4.3 1.3 red

6.1 2.7 4.3 1.3 red

5.5 2.4 3.7 1.0 red

5.8 2.7 3.9 1.2 red

6.0 2.7 5.1 1.6 blue

5.4 3.0 4.5 1.5 red

6.3 2.3 4.4 1.3 red

5.6 3.0 4.1 1.3 red

5.5 2.5 4.0 1.3 red

5.5 2.6 4.4 1.2 red

6.1 3.0 4.6 1.4 blue

5.0 2.3 3.3 1.0 red

5.6 2.7 4.2 1.3 red

5.7 3.0 4.2 1.2 red

5.7 2.9 4.2 1.3 red

5.1 2.5 3.0 1.1 red

6.0 3.4 4.5 1.6 blue

K = 3

Files: datasets/blue\_model\_8d.txt datasets/red\_model\_8d.txt datasets/unknown\_8d.txt

-1.252161 4.595982 0.403397 4.664438 1.184887 4.729431 0.483914 4.661279 blue

1.058662 4.139604 -0.743436 5.450736 -0.274353 4.145079 -0.467963 4.333563 blue

-0.835754 4.782662 -1.13831 5.45584 0.135823 4.216405 -0.231157 5.910545 blue

1.603099 4.092962 0.79424 5.514881 -1.259234 5.508506 -0.812622 4.289069 blue

-0.288341 4.550324 -0.597914 4.54299 -1.532651 5.006298 0.86868 4.734013 blue

-0.918091 5.803641 -0.406679 4.369684 -0.598748 5.700556 -0.294587 4.781506 blue

0.920304 4.52264 0.313416 4.17575 -1.23142 4.807567 0.873939 4.097733 blue

-1.892713 5.419602 0.145955 5.497594 -0.847049 4.472751 0.649144 5.086061 blue

0.990046 4.553275 0.740694 4.660983 -0.074056 4.290081 -0.216103 4.231506 blue

-0.623098 4.631027 0.470979 5.102245 1.139348 5.049803 1.562672 4.929757 blue

-1.917901 1.557697 1.61902 2.636847 0.345654 2.062426 0.32132 2.998103 red

-0.385049 3.148221 -0.72842 1.82968 -0.049924 2.331142 0.851898 2.402296 red

1.576337 2.411215 -0.884658 2.606353 -1.166365 2.880788 0.888455 1.649488 red

0.300287 1.865714 -0.168243 2.290775 1.491714 2.283442 -1.825572 2.360293 red

0.510346 1.917645 0.657498 1.74974 1.799944 2.373934 -1.568937 2.991486 red

-0.586001 2.052022 -0.665066 3.081557 0.885694 1.791796 1.468184 2.046193 red

1.818772 2.372645 -1.262745 2.019086 0.305039 2.753358 -0.487803 1.778518 red

1.358644 2.984491 -0.061581 1.98725 0.688312 1.813675 0.969963 2.196707 red

0.246576 2.359044 -1.153771 2.375196 0.141793 2.199259 0.52315 3.053531 red

0.374774 2.540291 -0.706522 2.680273 -1.404407 2.011466 1.878639 2.524433 red

K = 4

Files: datasets/blue\_model\_2d.txt datasets/red\_model\_2d.txt datasets/unknown\_2d.txt

0.678713 0.951598 blue

-1.169512 -0.957855 blue

0.631947 1.236489 blue

-0.131799 -0.324218 blue

-0.199565 -0.229171 blue

0.900907 0.448131 blue

0.699395 0.414314 blue

-1.132646 -1.013298 blue

1.004178 1.355361 blue

-0.127378 -0.097121 blue

6.08804 3.457729 red

4.147974 5.275341 red

6.538759 3.670323 red

4.579573 4.03559 red

4.756026 4.184762 red

5.221742 2.872705 red

5.271773 3.158064 red

4.046376 5.19232 red

6.530952 3.171413 red

4.918007 4.142507 red

K = 4

Files: datasets/blue\_model\_4d.txt datasets/red\_model\_4d.txt datasets/unknown\_4d.txt

5.5 2.4 3.8 1.1 red

5.7 2.8 4.1 1.3 red

5.8 2.6 4.0 1.2 red

6.2 2.9 4.3 1.3 blue

6.1 2.7 4.3 1.3 blue

5.5 2.4 3.7 1.0 red

5.8 2.7 3.9 1.2 red

6.0 2.7 5.1 1.6 blue

5.4 3.0 4.5 1.5 red

6.3 2.3 4.4 1.3 blue

5.6 3.0 4.1 1.3 red

5.5 2.5 4.0 1.3 red

5.5 2.6 4.4 1.2 red

6.1 3.0 4.6 1.4 blue

5.0 2.3 3.3 1.0 red

5.6 2.7 4.2 1.3 red

5.7 3.0 4.2 1.2 blue

5.7 2.9 4.2 1.3 blue

5.1 2.5 3.0 1.1 red

6.0 3.4 4.5 1.6 red

K = 4

Files: datasets/blue\_model\_8d.txt datasets/red\_model\_8d.txt datasets/unknown\_8d.txt

-1.252161 4.595982 0.403397 4.664438 1.184887 4.729431 0.483914 4.661279 blue

1.058662 4.139604 -0.743436 5.450736 -0.274353 4.145079 -0.467963 4.333563 blue

-0.835754 4.782662 -1.13831 5.45584 0.135823 4.216405 -0.231157 5.910545 blue

1.603099 4.092962 0.79424 5.514881 -1.259234 5.508506 -0.812622 4.289069 blue

-0.288341 4.550324 -0.597914 4.54299 -1.532651 5.006298 0.86868 4.734013 blue

-0.918091 5.803641 -0.406679 4.369684 -0.598748 5.700556 -0.294587 4.781506 blue

0.920304 4.52264 0.313416 4.17575 -1.23142 4.807567 0.873939 4.097733 blue

-1.892713 5.419602 0.145955 5.497594 -0.847049 4.472751 0.649144 5.086061 blue

0.990046 4.553275 0.740694 4.660983 -0.074056 4.290081 -0.216103 4.231506 blue

-0.623098 4.631027 0.470979 5.102245 1.139348 5.049803 1.562672 4.929757 blue

-1.917901 1.557697 1.61902 2.636847 0.345654 2.062426 0.32132 2.998103 red

-0.385049 3.148221 -0.72842 1.82968 -0.049924 2.331142 0.851898 2.402296 red

1.576337 2.411215 -0.884658 2.606353 -1.166365 2.880788 0.888455 1.649488 red

0.300287 1.865714 -0.168243 2.290775 1.491714 2.283442 -1.825572 2.360293 red

0.510346 1.917645 0.657498 1.74974 1.799944 2.373934 -1.568937 2.991486 red

-0.586001 2.052022 -0.665066 3.081557 0.885694 1.791796 1.468184 2.046193 red

1.818772 2.372645 -1.262745 2.019086 0.305039 2.753358 -0.487803 1.778518 red

1.358644 2.984491 -0.061581 1.98725 0.688312 1.813675 0.969963 2.196707 red

0.246576 2.359044 -1.153771 2.375196 0.141793 2.199259 0.52315 3.053531 red

0.374774 2.540291 -0.706522 2.680273 -1.404407 2.011466 1.878639 2.524433 red

Press any key to continue . . .