PROGRAMMING ASSIGNMENT 5 – REPORT

- The first 3 experiments use 2-d datapoints hence have been plotted using matplotlib for visual representation.

Experiment 1:

D = 2, q = 3

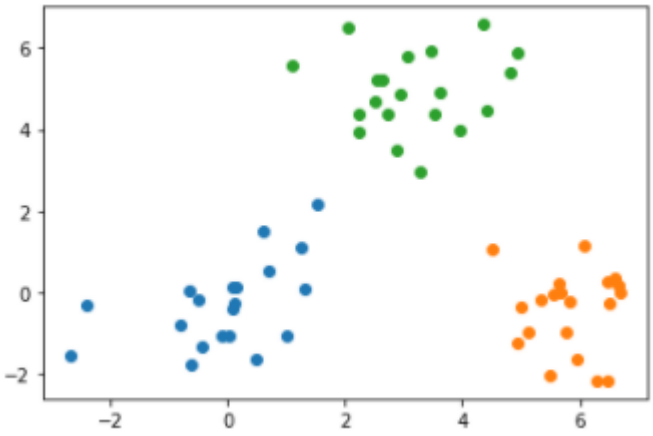
sigma = 1.0

mu1 = [0.0,0.0]

mu2 = [6.0,0.0]

mu3 = [3.0,3*math.sqrt(3)]

Randomly generated points:



1.1

K = 3, random restarts = 2

number of times kmeans returns less than k clusters: 0

average number of iterations: 4.0

Cost: 106.1559498690244

1.2

K = 3, random restarts = 5

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.0

Cost: 106.1559498690244

1.3

K = 3, random restarts = 10

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.5

Cost: 106.15594986902441

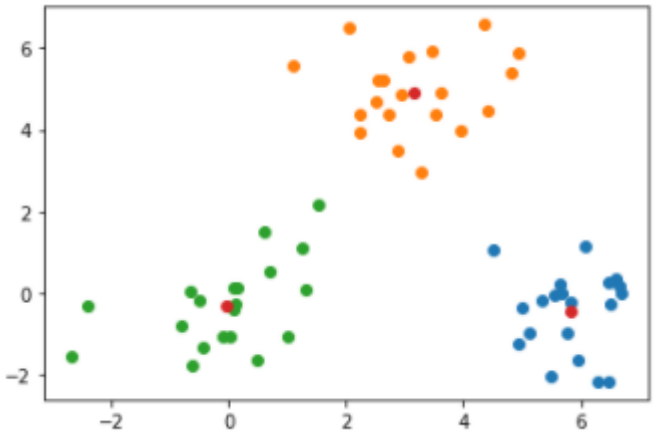
1.4

K = 3, random restarts = 100

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.56

Cost: 106.1559498690244



The convergence is almost instant with very low number of random restarts required.

Experiment 2:

D = 2, q =3

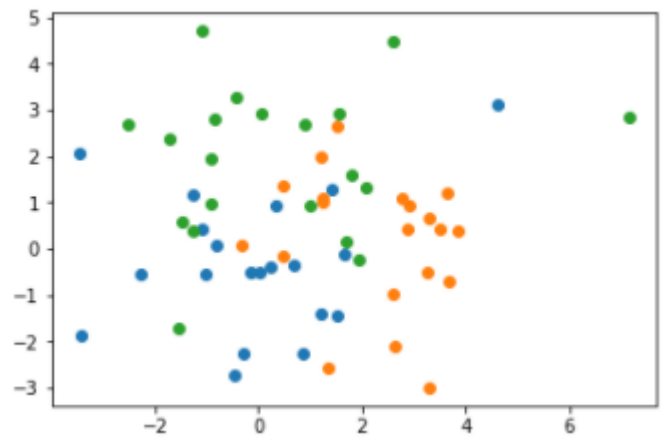
sigma = 2.0

mu1 = [0.0,0.0]

mu2 = [2.0,0.0]

mu3 = [1.0,math.sqrt(3)]

Randomly generated points:



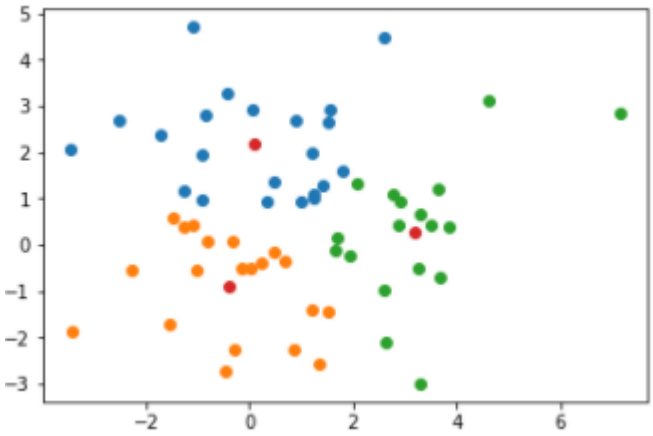
2.1

K = 3, random restarts = 2

number of times kmeans returns less than k clusters: 0

average number of iterations: 6.0

Cost: 192.1144950108524



2.2

K = 3, random restarts = 5

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.2

Cost: 191.73013064007392

2.3

K = 3, random restarts = 10

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.76

Cost: 191.73013064007392

2.4

K = 3, random restarts = 100

number of times kmeans returns less than k clusters: 0

average number of iterations: 4.44

Cost: 191.73013064007392

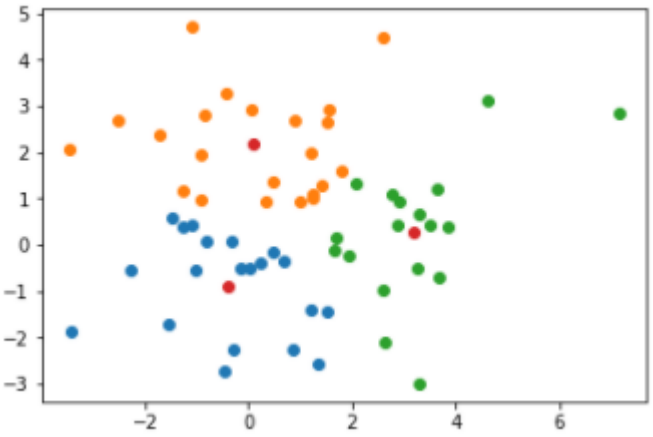
2.5

K = 3, random restarts = 1000

number of times kmeans returns less than k clusters: 0

average number of iterations: 4.359

Cost: 191.73013064007392



The convergence requires slightly higher number of random restarts

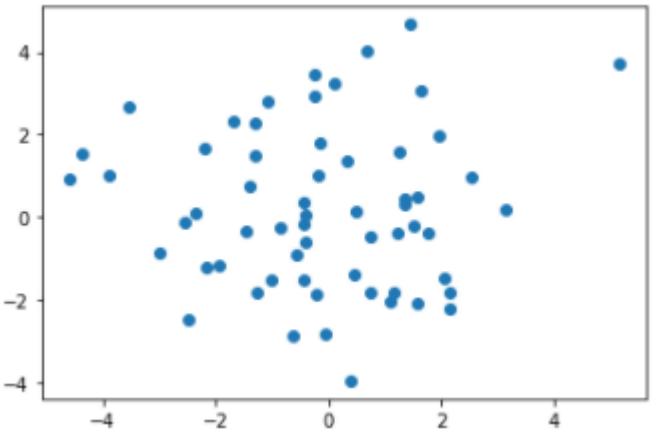
Experiment 3

D = 2, q = 1

sigma = 2.0

mu1 = [0.0,0.0]

Randomly generated points:



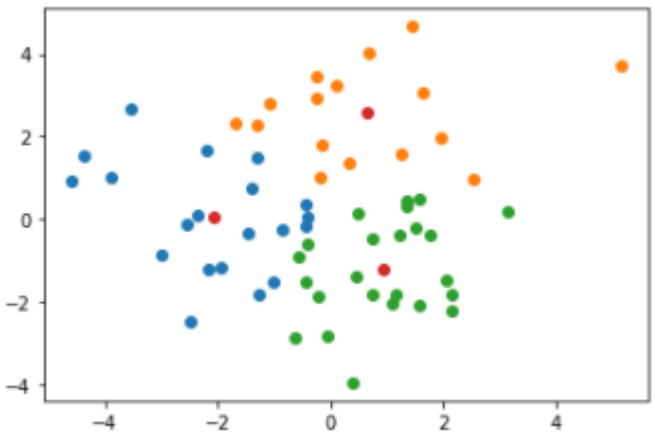
3.1

K = 3, random restarts = 2

number of times kmeans returns less than k clusters: 0

average number of iterations: 4.5

Cost: 179.58604495475407



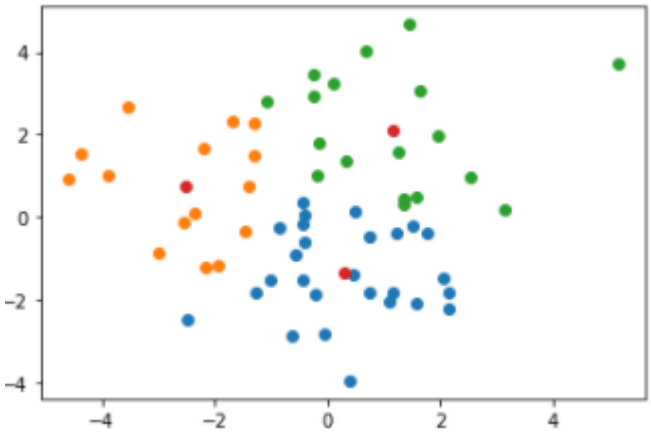
3.2

K = 3, random restarts = 5

number of times kmeans returns less than k clusters: 0

average number of iterations: 4.2

Cost: 177.4441697700281



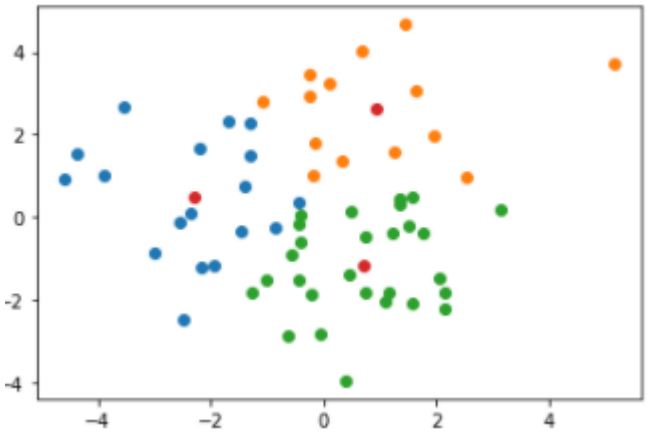
3.3

K = 3, random restarts = 10

number of times kmeans returns less than k clusters: 0

average number of iterations: 5.2

Cost: 175.95185077068922



3.4

K = 3, random restarts = 50

number of times kmeans returns less than k clusters: 0

average number of iterations: 5.62

Cost: 175.7145346643533

3.5

K = 3, random restarts = 100

number of times kmeans returns less than k clusters: 0

average number of iterations: 5.56

Cost: 175.7145346643533

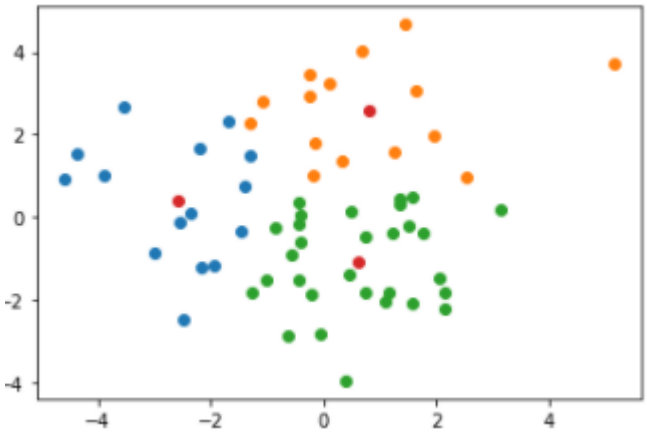
3.6

K = 3, random restarts = 1000

number of times kmeans returns less than k clusters: 0

average number of iterations: 5.444

Cost: 175.7145346643533



Experiment 4

D=100, q=3

sigma = 1.0

mu1 = [0.0]*100

mu2 = [60.0] + [0.0]*99

mu3 = [30.0, 30*math.sqrt(3)] + [0.0]*98

4.1

K = 3, random restarts = 2

number of times kmeans returns less than k clusters: 0

average number of iterations: 2.0

Cost: 5650.157762942369

4.2

K = 3, random restarts = 5

number of times kmeans returns less than k clusters: 0

average number of iterations: 2.2

Cost: 5650.157762942369

4.3

K = 3, random restarts = 10

number of times kmeans returns less than k clusters: 0

average number of iterations: 2.0

Cost: 5650.157762942369

4.4

K = 3, random restarts = 50

number of times kmeans returns less than k clusters: 0

average number of iterations: 2.18

Cost: 5650.157762942369

Experiment 5

D=100,q=3

sigma = 1.0

mu1 = [0.0]*100

mu2 = [6.0] + [0.0]*99

mu3 = [3.0, 3*math.sqrt(3)] + [0.0]*98

5.1

K = 3, random restarts = 2

number of times kmeans returns less than k clusters: 0

average number of iterations: 8.0

Cost: 5539.384844172206

5.2

K = 3, random restarts = 5

number of times kmeans returns less than k clusters: 0

average number of iterations: 7.0

Cost: 5531.911952365707

5.3

K = 3, random restarts = 10

number of times kmeans returns less than k clusters: 0

average number of iterations: 5.9

Cost: 5531.911952365707

5.4

K = 3, random restarts = 50

number of times kmeans returns less than k clusters: 0

average number of iterations: 8.48

Cost: 5531.911952365707

Experiment 6

D=100,q=3

sigma = 2.0

mu1 = [0.0]*100

mu2 = [2.0] + [0.0]*99

mu3 = [1.0, math.sqrt(3)] + [0.0]*98

6.1

K = 3, random restarts = 2

number of times kmeans returns less than k clusters: 0

average number of iterations: 2.5

Cost: 22684.289272170266

6.2

K = 3, random restarts = 5

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.2

Cost: 22685.332204847346

6.3

K = 3, random restarts = 10

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.5

Cost: 22570.15085289033

6.4

K = 3, random restarts = 50

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.42

Cost: 22533.127254589162

6.5

K = 3, random restarts = 100

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.32

Cost: 22532.82041068866

6.6

K = 3, random restarts = 150

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.1133333333333333

Cost: 22419.607586644426

6.7

3,200,0

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.465

Cost: 22508.705543637974

6.8

K = 3, random restarts = 1000

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.37

Cost: 22408.361170497465

6.9

K = 3, random restarts = 3000

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.387

Cost: 22366.659162215994

6.10

K = 3, random restarts = 5000

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.3763

Cost: 22406.48861034634

6.11

K = 3, random restarts = 10000

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.367

Cost: 22406.48861034634

Experiment 7

D=100,q=1

sigma = 2.0

mu1 = [0.0]*100

7.1

K = 3, random restarts = 2

number of times kmeans returns less than k clusters: 0

average number of iterations: 5.5

Cost: 22462.173577022128

7.2

K = 3, random restarts = 5

number of times kmeans returns less than k clusters: 0

average number of iterations: 4.0

Cost: 22380.531901784543

7.3

K = 3, random restarts = 10

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.6

Cost: 22386.090905192097

7.4

K = 3, random restarts = 50

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.26

Cost: 22426.606116931976

7.5

K = 3, random restarts = 100

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.25

Cost: 22369.864437926313

7.6

K = 3, random restarts = 1000

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.237

Cost: 22270.424080059878

7.7

K = 3, random restarts = 2000

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.272

Cost: 22268.189392557055

7.8

K = 3, random restarts = 10000

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.2945

Cost: 22281.998308935385

7.9

K = 3, random restarts = 50000

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.30556

Cost: 22223.257234009736

7.10

K = 3, random restarts = 100000

number of times kmeans returns less than k clusters: 0

average number of iterations: 3.29437

Cost: 22175.55755292663

Experiments show that experiment 7 takes the longest time to converge followed by experiment 6.