

# 수치해석

## *Final Term Project*

*Clustering of Vector Data*

2018007956 김채아

## [과정]

### < Clustering of vector data and classification >

#### ① Vectors are randomly generated with different Gaussian distributions

랜덤한 각 300개의 샘플들로 5개의 클러스터를 만든다

#### ② Use K-means Clustering

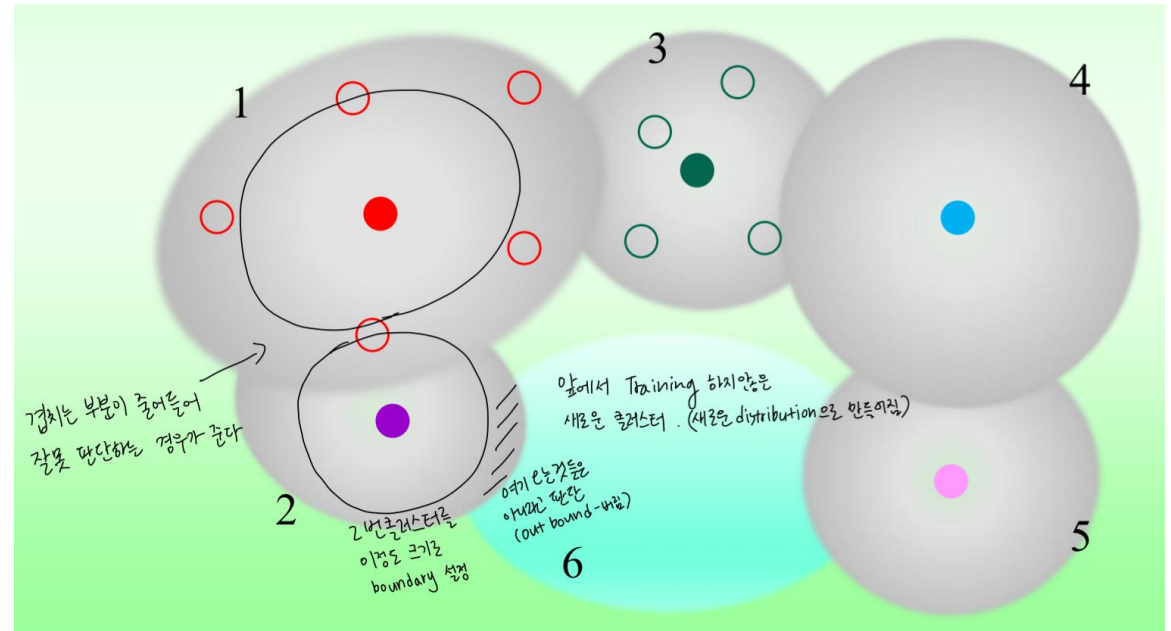
5개의 평균 벡터만 저장해둔다  
maximum distance에 대한 boundary를 설정하여  
샘플 300개 중 일부를 버린다  
(클러스터의 특성을 잘 나타내지 못하는 벡터들을 버림  
: mean vector에서 거리가 먼 벡터들)

#### ③ Test for randomly generated vectors

각 클러스터당 100개의 샘플들을 뽑아서  
해당되는 클러스터에 잘 클러스터링 되는지 확인한다

또한 5개의 클러스터에 해당하지 않는 새로운 샘플들 100개를 뽑아서  
5개의 클러스터 어디에도 속하지 않는다는 것을 보여준다

=> 위 과정을 두 번 해보았고, 각 경우 당 Testing을 두 번씩 해보았다



### [랜덤 숫자 생성]

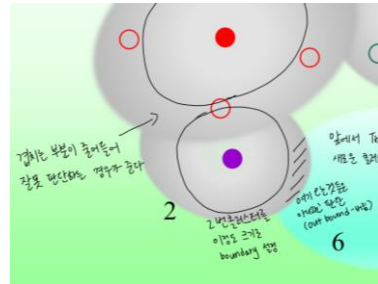
- 가우시안 분포를 적용할 중심 좌표를 1에서 50사이의 랜덤한 값으로 설정한다
- 가우시안 분포의 평균과 분산을 내가 직접 정해줘도 되지만, 이것도 랜덤하게 생성한다

⇒ 이 과정으로 구해진 300개 좌표들의 평균을 저장해둔다. 생성한 중심좌표, 가우시안 분포 값들도 저장해둔다

### [Maximum distance 설정]

범위를 설정하지 않고 최소거리에 해당하는 클러스터로 분류하게 되면 겹치는 부분이 많아서 잘못 클러스터링 되는 샘플들이 많이 생길 수 있다

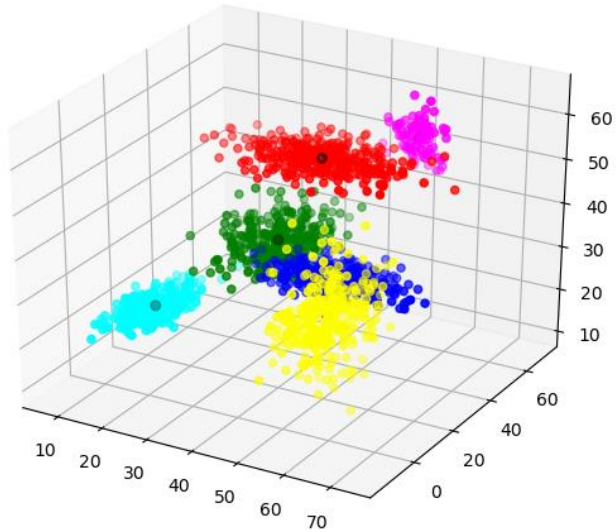
- > 각 클러스터마다 최대거리를 설정하고, 샘플이 해당 클러스터로 분류되더라도 최대 거리안에 들어오지 않으면 버린다
- 클러스터끼리 겹치는 범위를 줄여서 잘못 클러스터링 되는 경우를 줄인다



### [Testing]

- 저장해둔 중심좌표와 가우시안 분포값을 그대로 사용하여 각 클러스터당 100개를 랜덤하게 생성하고 각 클러스터의 mean vector와 비교해서 가까운 클러스터로 클러스터링 해본다
- 새로운 클러스터를 생성하여 클러스터 어디에도 속하지 않는다는 것을 보여준다  
새로운 클러스터는 미리 만들어진 중심 좌표들과 겹치지 않는 좌표를 잡고(50~55 범위에서 랜덤 생성), 평균, 표준편차를 작은 범위(1~5)의 랜덤한 숫자로 정한다  
(랜덤 생성 범위를 이렇게 잡은 이유는 다른 클러스터와 거의 섞이지 않는 경우를 만들어보기 위해서이고, 다른 클러스터와 같은 범위의 중심좌표, 가우시안 분포도 생성해서 테스트 해본다)

## (1) [Clustering of Vector Data]



### [클러스터의 평균벡터]

Red

mean: 36.54, 49.69, 49.01

Green

mean: 32.21, 37.43, 33.13

Blue

mean: 53.60, 17.97, 35.87

Yellow

mean: 55.05, 10.55, 29.97

Cyan

mean: 14.89, 14.68, 21.07

Magenta

mean: 54.80, 56.08, 55.74

각각 300개의 샘플을 가진 클러스터 5개를 생성  
(클러스터마다 100개씩의 샘플을 뽑아서 해당 클러스터로 클러스터링 되는지 Test)

0:red) right: 94 wrong: 2 out\_bound: 4  
mixed cluster: [0 1]

1:green) right: 95 wrong: 4 out\_bound: 1  
mixed cluster: [0 1 2]

2:blue) right: 80 wrong: 11 out\_bound: 9  
mixed cluster: [2 3]

3:yellow) right: 64 wrong: 29 out\_bound: 7  
mixed cluster: [2 3]

4:cyan) right: 100 wrong: 0 out\_bound: 0  
mixed cluster: [4]

100개의 샘플을 가진 새로운 클러스터를 생성  
(기존 클러스터 5개 중 어디에도 클러스터링 되지 않는다는 것을 Test)

new samples:magenta) wrong: 2 out\_bound: 98  
mixed cluster: [0]

max dist: [36.15834363258832, 39.28635605422784, 28.116981858453894,  
38.47128376478112, 35.28121231425473]

right: 원래의 클러스터로 잘 클러스터링 된 경우  
wrong: 다른 클러스터로 잘못 클러스터링 된 경우  
out\_bound: 클러스터 어디에도 속하지 못한 경우  
maximum distance: 각 클러스터의 최대 거리를  $x$ 라고 했을 때  
 $x-x/3$ , 즉 기존 클러스터의 약 2/3사이즈만큼으로  
범위를 설정했다

- 설정한 maximum distance의 문제점:  
x,y,z축으로의 분산이 다른데, 클러스터의 범위는 구 형태로 설정된다  
-> 가우시안 분포의 평균값이 작게 설정되면(평균벡터 근처로 많이 분포)  
문제가 없는데, 그렇지 않은 경우 wrong이 많이 발생할 것이다.  
하지만 분산에 따라 maximum distance를 타원 형태로 설정 하는 것을  
어떻게 구현 해야 할지 모르겠고 그렇게 하지않아도  
인식 정확도가 높게 나오고, 결과를 뽑는다는 무리가 없어서  
구 형태의 maximum distance를 사용했다

### [좌표들의 평균, 분산]

Red

X:(1,9) Y:(1,7) Z:(1,1)

Green

X:(2,5) Y:(9,9) Z:(1,3)

Blue

X:(6,7) Y:(3,5) Z:(8,1)

Yellow

X:(7,4) Y:(6,9) Z:(6,6)

Cyan

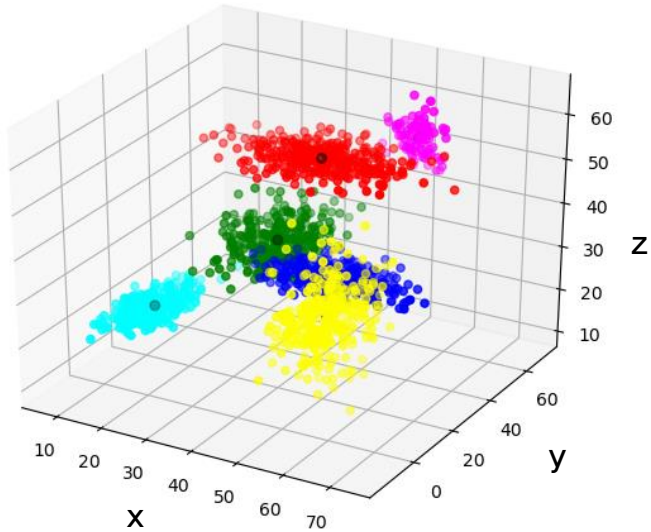
X:(4,3) Y:(4,8) Z:(9,1)

Magenta

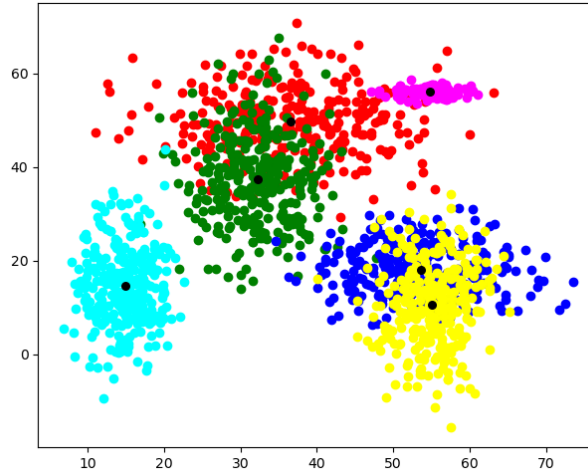
X:(3,3) Y:(4,1) Z:(3,3)

cluster number :  
R 0, G 1, B 2, Y 3, C 4

# (1) [Clustering of Vector Data 결과 분석]



x-y 평면 그래프



**Magenta** 자홍 클러스터는 조금 떨어진 거리에 완전 새로 뽑는다

mean: 54.80, 56.08, 55.74

Gauss: X:(3,3) Y:(4,1) Z:(3,3)

new samples) wrong: 3 out\_bound: 97

mixed cluster: [0]

자홍색 클러스터는 빨간색 클러스터와 살짝 겹치고 클러스터들 중에 가장 가까워서 빨간색 클러스터에 클러스터링 되지만 빨간색 클러스터에 maximum distance가 설정되어 있어서 거의 다 out\_bound로 분류된다  
인식 정확도 : 97%

**Red**

mean: 36.54, 49.69, 49.01

Gauss: X:(1,9) Y:(1,7) Z:(1,1)

max dist: 36.15

right: 92

wrong: 2

out\_bound: 6

mixed cluster: [0 1]

앞에서 말한 max\_dist의 한계 때문에 초록의 2개 샘플이 여기로 클러스터링 된다

x좌표의 분산이 커서 x축 방향의 끝부분은 out\_bound에 들어간다

인식 정확도 : 92%

**Green**

mean: 32.21, 37.43, 33.13

Gauss: X:(2,5) Y:(9,9) Z:(1,3)

max dist: 39.28

right: 66

wrong: 26

out\_bound: 8

mixed cluster: [0 1]

빨강과 초록이 x,y좌표가 겹쳐서 빨강이 초록으로 많이 분류된다

빨강 초록의 인식 정확도가 다른 이유는 각각의 분포형태와 설정된 max\_dist 때문이다

인식 정확도 : 66%

**Blue**

mean: 53.60, 17.97, 35.87

Gauss: X:(6,7) Y:(3,5) Z:(8,1)

max dist: 28.11

right: 45

wrong: 2

out\_bound: 53

mixed cluster: [2 3]

서로의 평균벡터가 비교적 가까운데 파랑이 x축방향으로 분산이 크고 노랑은 y축방향으로 분산이 커서 겹치는 부분이 많이 발생한다 인식 정확도가 떨어진다

인식 정확도 : 45%

타원 모양의 분포가 찌그러질수록 설정한 max\_dist의 특성상(구 모양) out\_bound가 더 생기게 된다

**Yellow**

mean: 55.05, 10.55, 29.97

Gauss: X:(7,4) Y:(6,9) Z:(6,6)

max dist: 38.47

right: 39

wrong: 33

out\_bound: 28

mixed cluster: [1 2 3]

인식 정확도 : 39%

**Cyan**

mean: 14.89, 14.68, 21.07

Gauss: X:(4,3) Y:(4,8) Z:(9,1)

max dist: 35.28

right: 76

wrong: 13

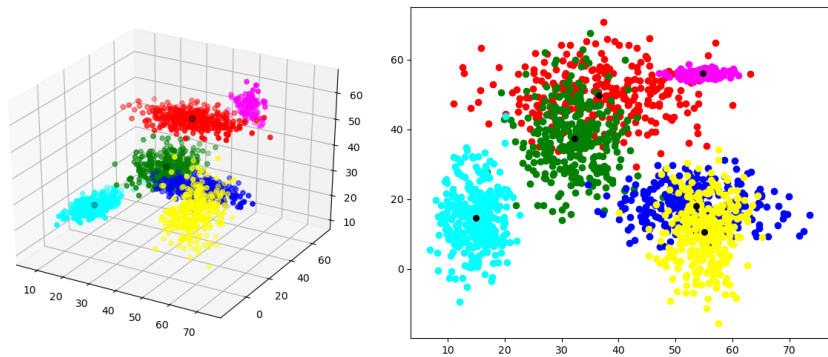
out\_bound: 11

mixed cluster: [1 4]

y축방향으로 분산이 커서 끝부분 샘플들은 청록보다 초록에 가깝게 분류된다

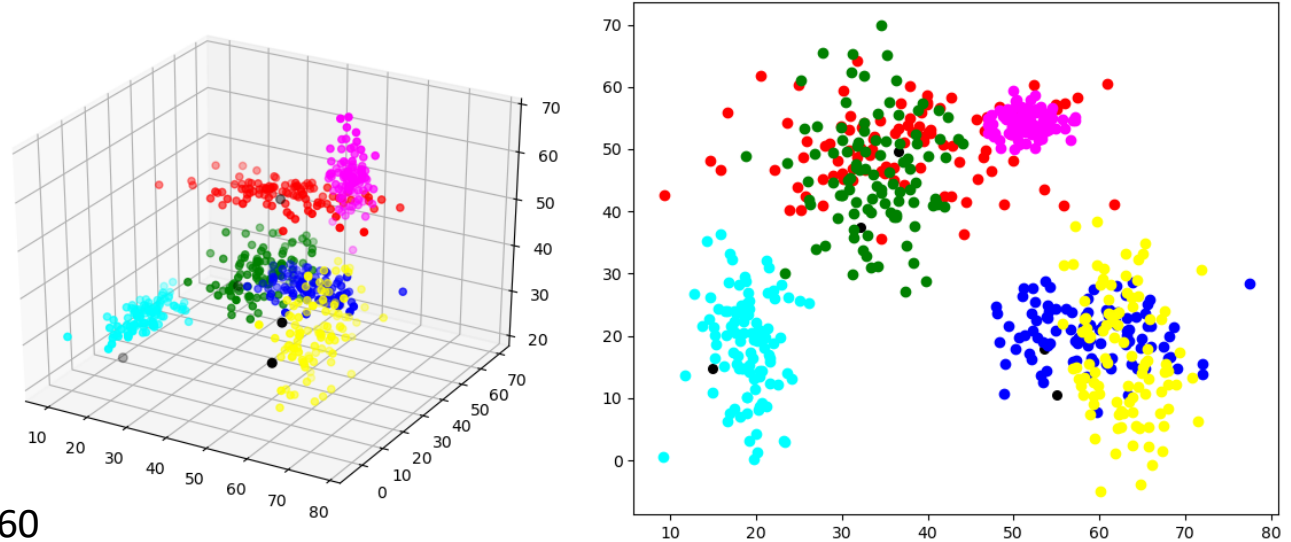
인식 정확도 : 76%

# (1) [Testing - ①]



기존 분포

## 각 클러스터에서 100개씩 추출 & 새로운 자홍 클러스터 구성



### Magenta

mean: 51.59, 54.26, 54.94  
Gauss: X:(2,2) Y:(3,2) Z:(4,4)

new samples) wrong: 40 out\_bound: 60  
mixed cluster: [0] 인식 정확도 : 60%  
빨강이 x축으로의 분포가 커서  
maximum distance안에 자홍이 들어가는 경우가  
많이 생긴다

왼쪽 그래프의 검은 점은 클러스터의 평균벡터를 나타낸 것인데  
오른쪽 그래프의 검은 점은 왼쪽 그래프에 있는 평균벡터  
즉 원래 클러스터의 평균벡터를 나타낸 것이다  
평균과 분산에 따라 새로 추출된 좌표들이 분포하는 모양을 보여준다

### Red

mean: 36.54, 49.69, 49.01  
Gauss: X:(1,9) Y:(1,7) Z:(1,1)  
max dist: 36.15  
right: 88  
wrong: 0  
out\_bound: 12  
mixed cluster: [0]

### Green

mean: 32.21, 37.43, 33.13  
Gauss: X:(2,5) Y:(9,9) Z:(1,3)  
max dist: 39.28  
right: 72  
wrong: 15  
out\_bound: 13  
mixed cluster: [0 1]

### Blue

mean: 53.60, 17.97, 35.87  
Gauss: X:(6,7) Y:(3,5) Z:(8,1)  
max dist: 28.11  
right: 53  
wrong: 2  
out\_bound: 45  
mixed cluster: [2 3]

### Yellow

mean: 55.05, 10.55, 29.97  
Gauss: X:(7,4) Y:(6,9) Z:(6,6)  
max dist: 38.47  
right: 33  
wrong: 28  
out\_bound: 39  
mixed cluster: [2 3]

### Cyan

mean: 14.89, 14.68, 21.07  
Gauss: X:(4,3) Y:(4,8) Z:(9,1)  
max dist: 35.28  
right: 74  
wrong: 18  
out\_bound: 8  
mixed cluster: [1 4]

인식 정확도 : 88%

인식 정확도 : 72%

인식 정확도 : 53%

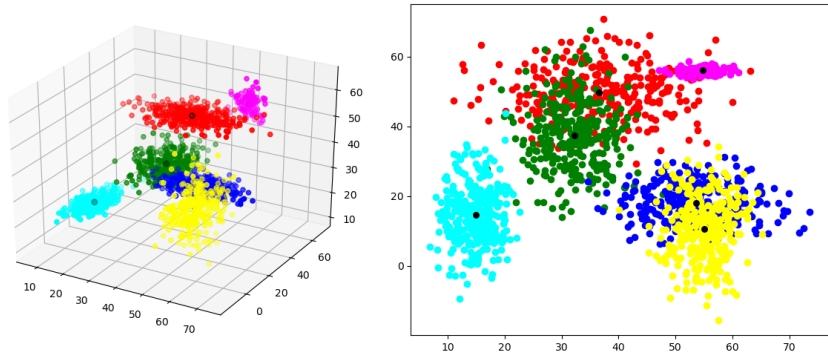
인식 정확도 : 33%

인식 정확도 : 74%

가우시안 분포의 평균값이 기존 클러스터의 평균벡터와 멀고, 분산이 크게 설정되면 out\_bound가 많이 나온다  
노랑 클러스터가 다른 클러스터로 잘못 클러스터링 되는 경우가 가장 많다

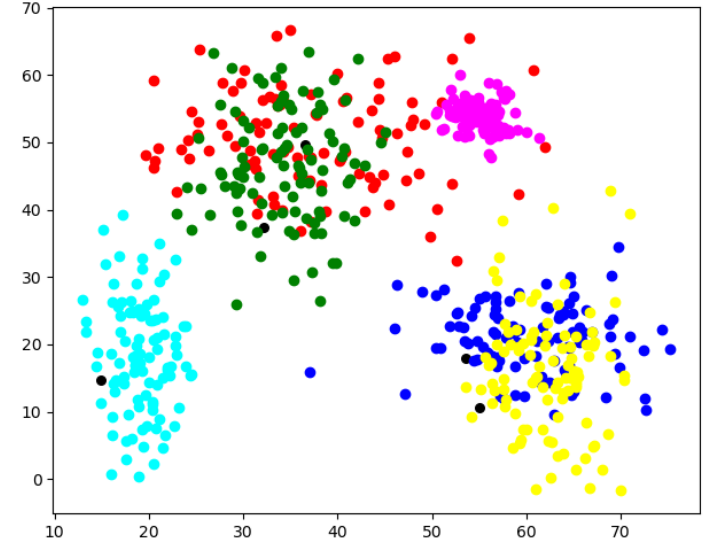
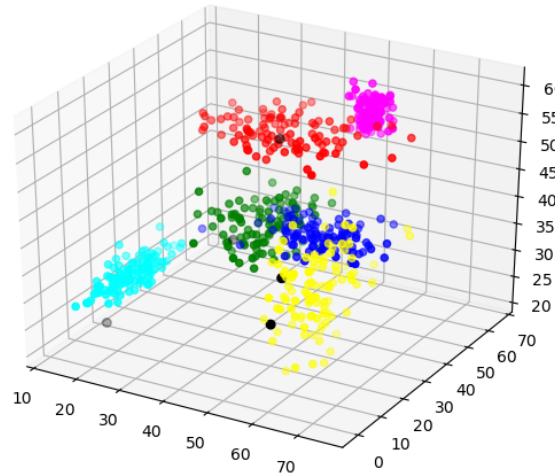


# (1) [Testing – ②]



기존 분포

각 클러스터에서 100개씩 추출 & 새로운 자홍 클러스터 구성



자홍이 빨강과 많이 겹쳐지는 것으로 보이는데  
빨강의 maximum\_distance에 걸려서 클러스터링 되지않고  
대부분 out\_bound로 잘 분류된다

## Magenta

mean: 54.99, 54.06, 55.73  
Gauss: X:(3,2) Y:(2,2) Z:(1,2)  
new samples) wrong: 4 out\_bound: 96  
mixed cluster: [0] 인식 정확도 : 96%

## Red

mean: 36.54, 49.69, 49.01  
Gauss: X:(1,9) Y:(1,7) Z:(1,1)  
max dist: 36.15  
right: 89  
wrong: 0  
out\_bound: 11  
mixed cluster: [0 2]

인식 정확도 : 89%

## Green

mean: 32.21, 37.43, 33.13  
Gauss: X:(2,5) Y:(9,9) Z:(1,3)  
max dist: 39.28  
right: 74  
wrong: 16  
out\_bound: 10  
mixed cluster: [0 1]

인식 정확도 : 74%

## Blue

mean: 53.60, 17.97, 35.87  
Gauss: X:(6,7) Y:(3,5) Z:(8,1)  
max dist: 28.11  
right: 41  
wrong: 2  
out\_bound: 57  
mixed cluster: [2 3]

인식 정확도 : 41%

## Yellow

mean: 55.05, 10.55, 29.97  
Gauss: X:(7,4) Y:(6,9) Z:(6,6)  
max dist: 38.47  
right: 32  
wrong: 39  
out\_bound: 29  
mixed cluster: [1 2 3]

인식 정확도 : 32%

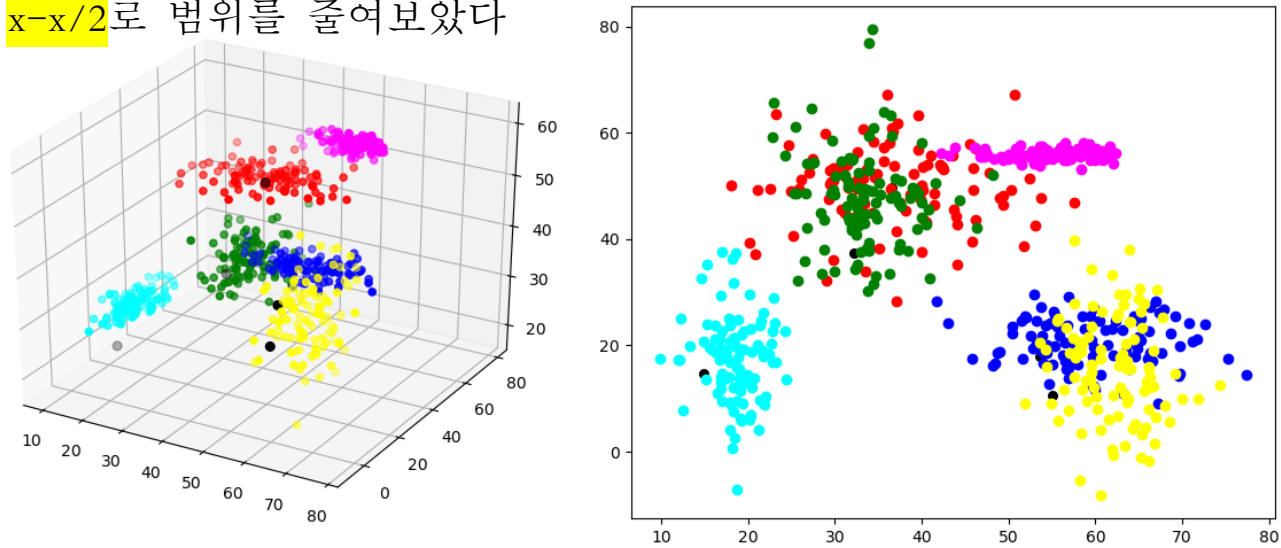
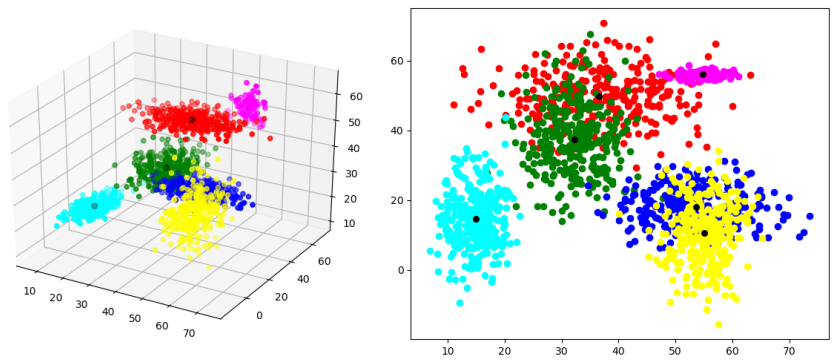
## Cyan

mean: 14.89, 14.68, 21.07  
Gauss: X:(4,3) Y:(4,8) Z:(9,1)  
max dist: 35.28  
right: 75  
wrong: 15  
out\_bound: 10  
mixed cluster: [1 4]

인식 정확도 : 75%

(1) [클러스터의 maximum distance 범위 줄여봄]

이전 테스트에서는 각 클러스터의 최대 거리를  $x$ 라고 했을 때  $x-x/3$ , 즉 기존 클러스터의 약 2/3사이즈만큼으로 범위를 설정했는데  $x-x/2$ 로 범위를 줄여보았다



Magenta  
mean: 54.50, 55.93, 58.05  
Gauss: X:(3,4) Y:(3,1) Z:(4,1)

new samples) wrong: 0 out\_bound: 100  
mixed cluster: [0] 인식 정확도 : 100%

자홍은 모두 out\_bound로 분류되었다

Red  
mean: 36.54, 49.69, 49.01  
Gauss: X:(1,9) Y:(1,7) Z:(1,1)  
max dist: 36.15  
right: 78  
wrong: 0  
out\_bound: 22  
mixed cluster: [0 1]  
인식 정확도 : 78%

Green  
mean: 32.21, 37.43, 33.13  
Gauss: X:(2,5) Y:(9,9) Z:(1,3)  
max dist: 39.28  
right: 63  
wrong: 10  
out\_bound: 27  
mixed cluster: [0 1]  
인식 정확도 : 63%

Blue  
mean: 53.60, 17.97, 35.87  
Gauss: X:(6,7) Y:(3,5) Z:(8,1)  
max dist: 28.11  
right: 19  
wrong: 1  
out\_bound: 80  
mixed cluster: [1 2 3]  
인식 정확도 : 19%

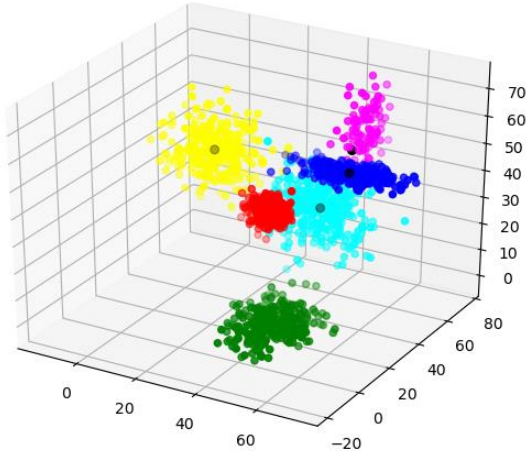
Yellow  
mean: 55.05, 10.55, 29.97  
Gauss: X:(7,4) Y:(6,9) Z:(6,6)  
max dist: 38.47  
right: 29  
wrong: 17  
out\_bound: 54  
mixed cluster: [2 3]  
인식 정확도 : 29%

Cyan  
mean: 14.89, 14.68, 21.07  
Gauss: X:(4,3) Y:(4,8) Z:(9,1)  
max dist: 35.28  
right: 48  
wrong: 4  
out\_bound: 48  
mixed cluster: [1 4]  
인식 정확도 : 48%

모든 클러스터의 out\_bound 비율이 커졌고, 잘못 클러스터링 되는 경우인 wrong이 작아졌다  
인식 비율이 줄어들지만, 잘못 클러스터링 되는 오류를 줄일 수 있다



(2) [또 다른 Clustering of Vector Data] 각각 300개의 샘플을 가진 클러스터 5개를 생성  
(클러스터마다 100개씩의 샘플을 뽑아서 해당 클러스터로 클러스터링 되는지 Test)



[클러스터의 평균벡터]

Red

mean: 44.92, 9.97, 44.97

Green

mean: 49.24, 4.20, 7.16

Blue

mean: 53.88, 43.21, 47.96

Yellow

mean: 6.35, 48.01, 43.25

Cyan

mean: 41.77, 48.73, 29.82

Magenta

mean: 50.65, 59.84, 57.34

0:red) right: 45 wrong: 0 out\_bound: 55

mixed cluster: [0]

1:green) right: 87 wrong: 0 out\_bound: 13

mixed cluster: [1]

2:blue) right: 76 wrong: 0 out\_bound: 24

mixed cluster: [2]

3:yellow) right: 57 wrong: 0 out\_bound: 43

mixed cluster: [3]

4:cyan) right: 49 wrong: 38 out\_bound: 13

mixed cluster: [2 3 4]

100개의 샘플을 가진 새로운 클러스터를 생성

(기존 클러스터 5개 중 어디에도 클러스터링 되지 않는다는 것을 Test)

new samples:magenta) wrong: 13 out\_bound: 87

mixed cluster: [2 4]

max dist: [16.251207675436415, 38.29365190417902, 31.328264644590796, 35.9345701007087, 50.09883835501696]

[좌표들의 평균, 분산]

Red

X:(4,3) Y:(3,2) Z:(5,3)

Green

X:(8,6) Y:(3,9) Z:(2,3)

Blue

X:(7,8) Y:(5,4) Z:(2,1)

Yellow

X:(3,7) Y:(8,6) Z:(7,7)

Cyan

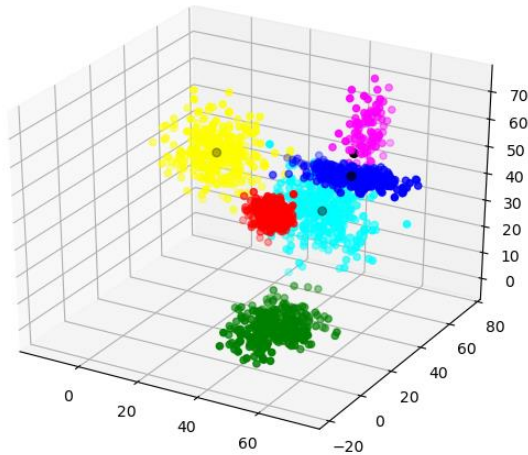
X:(6,8) Y:(3,6) Z:(8,8)

Magenta

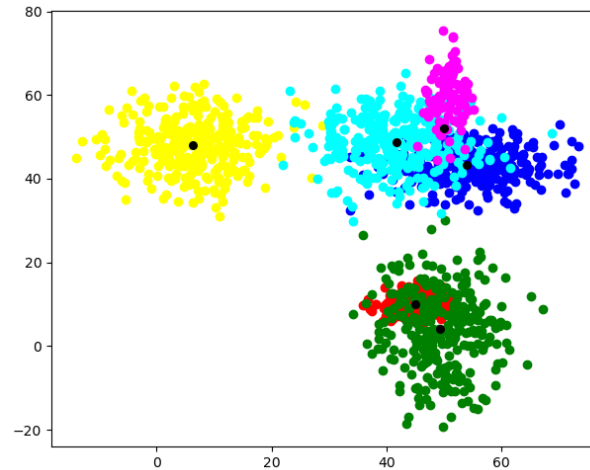
X:(1,2) Y:(8,6) Z:(6,7)

(자홍 클러스터의 분산을 조금 키웠다)

## (2) [Clustering of Vector Data 결과 분석]



x-y 평면 그래프



### Magenta

mean: 50.65, 59.84, 57.34

Gauss: X:(1,2) Y:(8,6) Z:(6,7)

new samples) wrong: 13 out\_bound: 87

mixed cluster: [2 4]

자홍색 클러스터는 파랑, 청록에 클러스터링 되지만  
두 클러스터 각각의 maximum distance에 의해  
대부분 out\_bound로 분류된다

인식 정확도 : 87%

### Red

mean: 44.92, 9.97, 44.97

Gauss: (4,3) Y:(3,2) Z:(5,3)

max dist: 16.25

right: 45

wrong: 0

out\_bound: 55

mixed cluster: [0]

다른 클러스터로 잘못 분류 되지  
않고, 샘플과 mean과의 거리가  
maximum distance  
이상이면 out\_bound로 분류된다

인식 정확도 : 45%

### Green

mean: 49.24, 4.20, 7.16

Gauss: X:(8,6) Y:(3,9) Z:(2,3)

max dist: 38.29

right: 87

wrong: 0

out\_bound: 13

mixed cluster: [1]

다른 클러스터로 잘못 분류 되지 않고  
샘플과 mean과의 거리가  
maximum distance이상이면  
out\_bound로 분류된다

인식 정확도 : 87%

### Blue

mean: 53.88, 43.21, 47.96

Gauss: X:(7,8) Y:(5,4) Z:(2,1)

max dist: 31.32

right: 76

wrong: 0

out\_bound: 24

mixed cluster: [2]

청록에 비해 분산이 작아서  
청록으로 분류될 것들이 다  
out\_bound에 속해서 분류된다

(자홍은 나중에 만들어지는  
새로운 클러스터이므로  
여기서는 생각하지 않는다)

인식 정확도 : 76%

### Yellow

mean: 6.35, 48.01, 43.25

Gauss: X:(3,7) Y:(8,6) Z:(7,7)

max dist: 35.93

right: 57

wrong: 0

out\_bound: 43

mixed cluster: [3]

빨강, 초록 클러스터처럼  
본인으로만 클러스터링 된다

인식 정확도 : 57%

### Cyan

mean: 41.77, 48.73, 29.82

Gauss: X:(6,8) Y:(3,6) Z:(8,8)

max dist: 50.09

right: 49

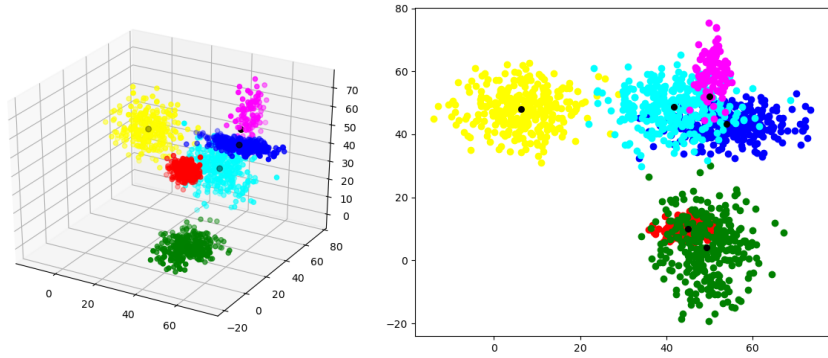
wrong: 38

out\_bound: 13

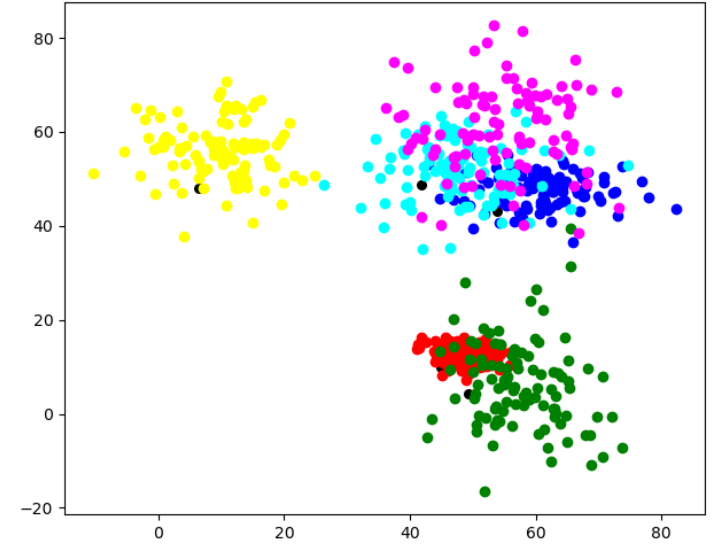
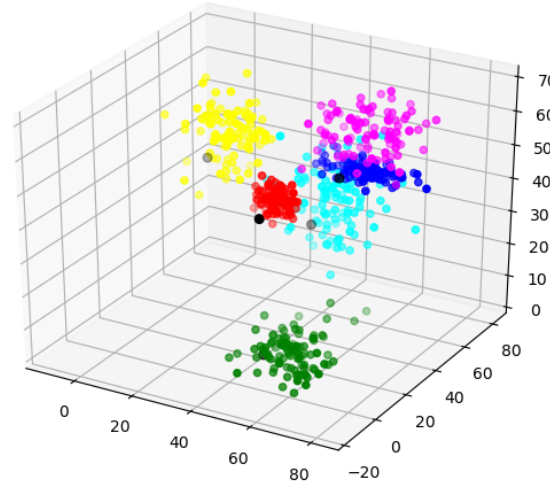
mixed cluster: [2 3 4]

파랑으로 47개,  
노랑으로 1개,  
청록으로 52개 분류된다  
이 중 out\_bound되는 것을 빼버리면  
위와 같은 결과가 나온다  
파랑으로 잘못 분류되는 경우가 많다  
인식 정확도 : 49%

## (2) [Testing - ①] 클러스터마다 100개씩의 샘플을 뽑고, 새로운 해당 클러스터로 클러스터링 되는지 Test



기존 분포



### Magenta

mean: 54.56, 60.96, 56.11      new samples) wrong: 11 out\_bound: 89  
 Gauss: X:(1,9) Y:(6,9) Z:(3,5)      mixed cluster: [2]      인식 정확도 : 89%

### Red

mean: 44.92, 9.97, 44.97  
 Gauss: (4,3) Y:(3,2) Z:(5,3)  
 max dist: 16.25  
 right: 40  
 wrong: 0  
 out\_bound: 60  
 mixed cluster: [0]

인식 정확도 : 50%

### Green

mean: 49.24, 4.20, 7.16  
 Gauss: X:(8,6) Y:(3,9) Z:(2,3)  
 max dist: 38.29  
 right: 80  
 wrong: 0  
 out\_bound: 20  
 mixed cluster: [1 4]

인식 정확도 : 87%

### Blue

mean: 53.88, 43.21, 47.96  
 Gauss: X:(7,8) Y:(5,4) Z:(2,1)  
 max dist: 31.32  
 right: 82  
 wrong: 0  
 out\_bound: 18  
 mixed cluster: [2]

인식 정확도 : 83%

### Yellow

mean: 6.35, 48.01, 43.25  
 Gauss: X:(3,7) Y:(8,6) Z:(7,7)  
 max dist: 35.93  
 right: 64  
 wrong: 0  
 out\_bound: 36  
 mixed cluster: [3]

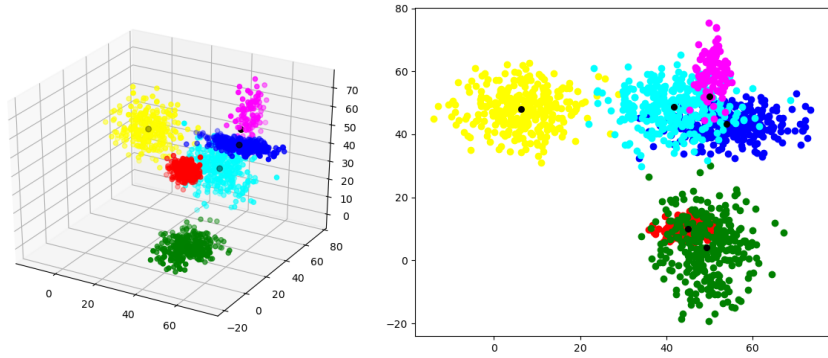
인식 정확도 : 69%

### Cyan

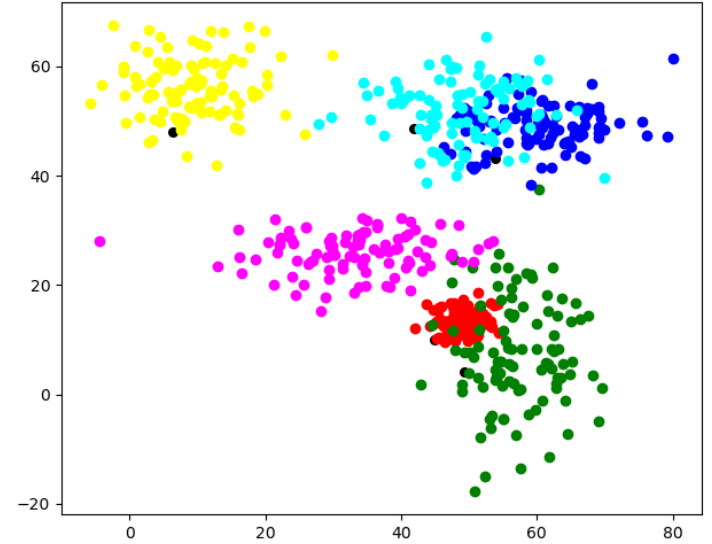
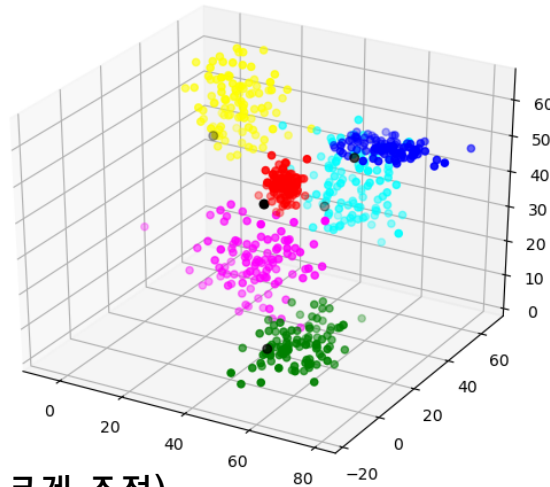
mean: 41.77, 48.73, 29.82  
 Gauss: X:(6,8) Y:(3,6) Z:(8,8)  
 max dist: 50.09  
 right: 58  
 wrong: 21  
 out\_bound: 21  
 mixed cluster: [2 3 4]

인식 정확도 : 67%

## (2) [Testing – ②]



기존 분포



자홍의 랜덤한 범위를 다르게 설정함

**Magenta** (평균벡터들 사이에 중심좌표를 잡고, 가우시안 평균, 분산은 더 크게 조절)

mean: 33.18, 25.99, 20.33      new samples) wrong: 22 out\_bound: 78

Gauss: X:(6,9) Y:(5,4) Z:(5,8)      mixed cluster: [0 1 2 3 4]      인식 정확도 : 78%

**Red**

mean: 44.92, 9.97, 44.97

Gauss: (4,3) Y:(3,2) Z:(5,3)

max dist: 16.25

right: 41      인식 정확도 : 41%

wrong: 0

out\_bound: 59

mixed cluster: [0]

빨강 클러스터가 뭉쳐있지만 out\_bound로 많이 분류되는 이유 :

분산이 작아서 maximum distance가 작게 설정되는데 비해

가우시안 평균이 크게 설정되어 샘플들이 평균벡터로부터 좀 떨어져서 분포하게된다

그래서 많은 샘플들이 maximum distance보다 멀어져서 out\_bound로 분류된다

**Green**

mean: 49.24, 4.20, 7.16

Gauss: X:(8,6) Y:(3,9) Z:(2,3)

max dist: 38.29

right: 79      인식 정확도 : 79%

wrong: 0

out\_bound: 21

mixed cluster: [1]

**Blue**

mean: 53.88, 43.21, 47.96

Gauss: X:(7,8) Y:(5,4) Z:(2,1)

max dist: 31.32

right: 80      인식 정확도 : 80%

wrong: 0

out\_bound: 20

mixed cluster: [2]

**Yellow**

mean: 6.35, 48.01, 43.25

Gauss: X:(3,7) Y:(8,6) Z:(7,7)

max dist: 35.93

right: 63

wrong: 0

out\_bound: 37

mixed cluster: [3]

인식 정확도 : 63%

**Cyan**

mean: 41.77, 48.73, 29.82

Gauss: X:(6,8) Y:(3,6) Z:(8,8)

max dist: 50.09

right: 65

wrong: 18

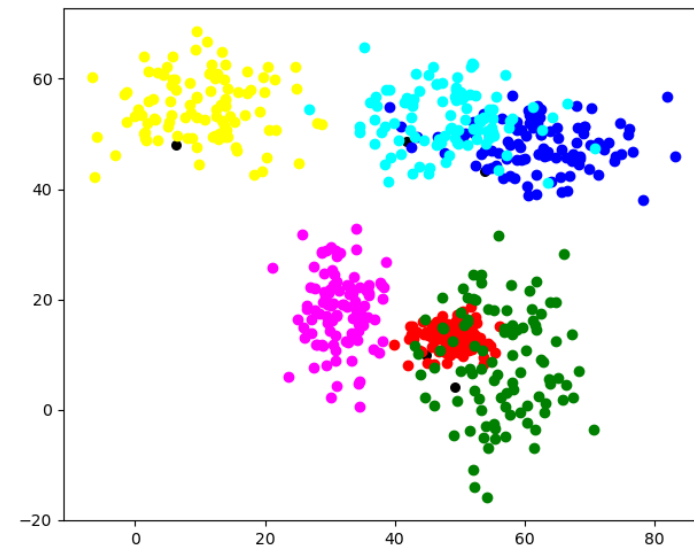
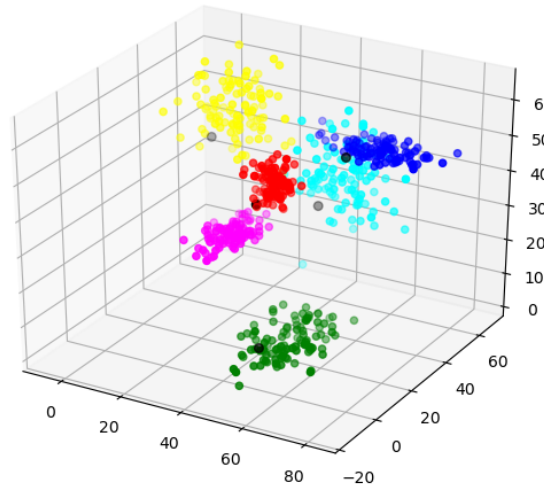
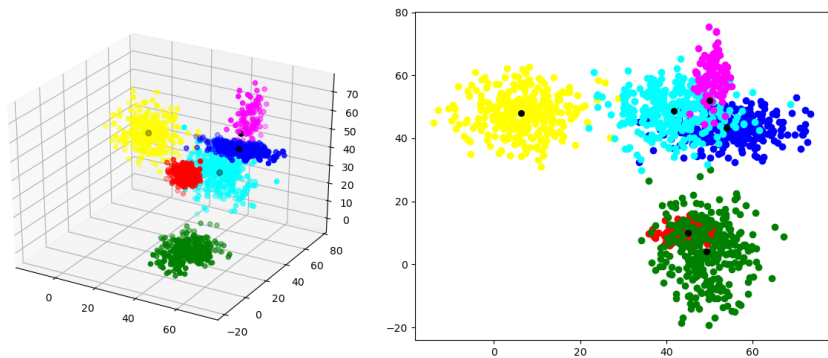
out\_bound: 17

mixed cluster: [2 3 4]

인식 정확도 : 65%

## (2) [클러스터의 maximum distance 범위 줄여봄]

이전 테스트에서는 각 클러스터의 최대 거리를  $x$ 라고 했을 때  $x-x/3$ , 즉 기존 클러스터의 약 2/3사이즈만큼으로 범위를 설정했는데  $x-x/2$ 로 범위를 줄여보았다



### Magenta

mean: 31.60, 18.15, 31.02    new samples) wrong: 0 out\_bound: 100  
Gauss: X:(2,4) Y:(4,7) Z:(1,1)    mixed cluster: [0 4]    인식 정확도 : 100%

### Red

mean: 36.54, 49.69, 49.01  
Gauss: X:(1,9) Y:(1,7) Z:(1,1)  
max dist: 36.15  
right: 18  
wrong: 0  
out\_bound: 82  
mixed cluster: [0]

인식 정확도 : 18%

### Green

mean: 32.21, 37.43, 33.13  
Gauss: X:(2,5) Y:(9,9) Z:(1,3)  
max dist: 39.28  
right: 50  
wrong: 0  
out\_bound: 50  
mixed cluster: [1]

인식 정확도 : 50%

### Blue

mean: 53.60, 17.97, 35.87  
Gauss: X:(6,7) Y:(3,5) Z:(8,1)  
max dist: 28.11  
right: 48  
wrong: 0  
out\_bound: 52  
mixed cluster: [2]

인식 정확도 : 48%

### Yellow

mean: 55.05, 10.55, 29.97  
Gauss: X:(7,4) Y:(6,9) Z:(6,6)  
max dist: 38.47  
right: 37  
wrong: 1  
out\_bound: 62  
mixed cluster: [3 4]

인식 정확도 : 37%

### Cyan

mean: 14.89, 14.68, 21.07  
Gauss: X:(4,3) Y:(4,8) Z:(9,1)  
max dist: 35.28  
right: 49  
wrong: 15  
out\_bound: 36  
mixed cluster: [2 4]

인식 정확도 : 49%

모든 클러스터의 out\_bound 비율이 커졌다

(이 분포는 원래 wrong되는게 거의 없었으므로 자홍 클러스터를 생각해보면) 자홍을 모두 out\_bound로 인식하는 것을 볼 수 있다