

수치해석 Final

Numerical analysis

JaeHoon KANG

강 재 훈

Final

CONTENTS



01 준비 과정



02 실행 결과



03 생각 해볼 점

$$\text{eq1: } x^2 + y^2 + z^2 = 1$$

$$\text{eq2: } (x-1)^2 + (y+1)^2 + z^2 / 3 = 1$$

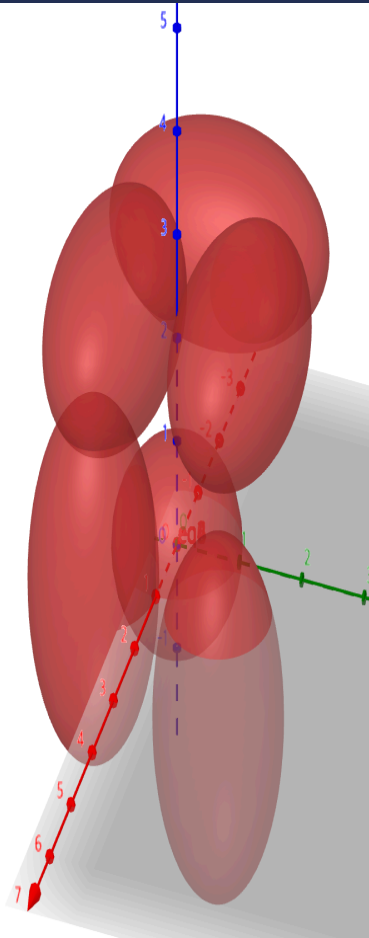
$$\text{eq3: } x^2 / 3 + (y+1)^2 + (z-2)^2 = 1$$

$$\text{eq4: } (x+2)^2 + \frac{y^2}{3} + (z-2)^2 = 1$$

$$\text{eq5: } \frac{x^2}{3} + (y-1)^2 + (z-2)^2 = 1$$

$$\text{eq6: } (x-1)^2 + (y-1)^2 + (z+1)^2 / 3 = 1$$

입력...



Distribution 정하기

서로 다른 6개의 x, y, z 좌표 분포를 generate합니다.
따라서 x, y, z 각각에
적당한 가우시안 distribution을 적용합니다.

$$x_1, y_1, z_1 = N(0, 1), N(0, 1), N(0, 1)$$

$$x_2, y_2, z_2 = N(1, 1^{**}0.5), N(-1, 1^{**}0.5), N(0, 3^{**}0.5)$$

$$x_3, y_3, z_3 = N(0, 3^{**}0.5), N(-1, 1^{**}0.5), N(2, 1^{**}0.5)$$

$$x_4, y_4, z_4 = N(-2, 1^{**}0.5), N(0, 3^{**}0.5), N(2, 1^{**}0.5)$$

$$x_5, y_5, z_5 = N(0, 3^{**}0.5), N(1, 1^{**}0.5), N(2, 1^{**}0.5)$$

$$x_6, y_6, z_6 = N(1, 1^{**}0.5), N(1, 1^{**}0.5), N(-1, 3^{**}0.5)$$

$$\text{eq1: } x^2 + y^2 + z^2 = 1$$

$$\text{eq2: } (x-1)^2 + (y+1)^2 + z^2 / 3 = 1$$

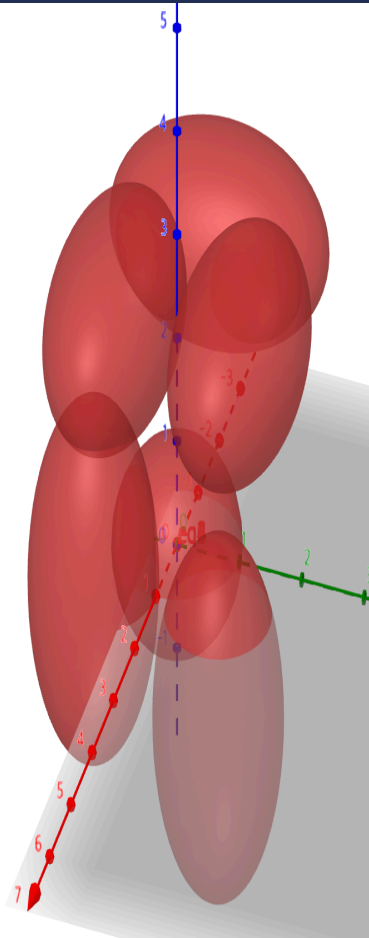
$$\text{eq3: } x^2 / 3 + (y+1)^2 + (z-2)^2 = 1$$

$$\text{eq4: } (x+2)^2 + \frac{y^2}{3} + (z-2)^2 = 1$$

$$\text{eq5: } \frac{x^2}{3} + (y-1)^2 + (z-2)^2 = 1$$

$$\text{eq6: } (x-1)^2 + (y-1)^2 + (z+1)^2 / 3 = 1$$

입력...



Distribution 정하기

$$x_1, y_1, z_1 = N(0, 1), N(0, 1), N(0, 1)$$

$$x_2, y_2, z_2 = N(1, 1^{**0.5}), N(-1, 1^{**0.5}), N(0, 3^{**0.5})$$

$$x_3, y_3, z_3 = N(0, 3^{**0.5}), N(-1, 1^{**0.5}), N(2, 1^{**0.5})$$

$$x_4, y_4, z_4 = N(-2, 1^{**0.5}), N(0, 3^{**0.5}), N(2, 1^{**0.5})$$

$$x_5, y_5, z_5 = N(0, 3^{**0.5}), N(1, 1^{**0.5}), N(2, 1^{**0.5})$$

$$x_6, y_6, z_6 = N(1, 1^{**0.5}), N(1, 1^{**0.5}), N(-1, 3^{**0.5})$$

1번부터 5번 분포까지는 data를 generate 시켜
K-clustering을 하기 위해 300개씩 sample을 뽑습니다.

이후 1번부터 6번 분포까지 각각에서
새로운 100개의 sample을 뽑아
각각의 sample들이 어느 cluster에 들어가는지
확인하고 결과를 분석하였습니다.

```
center :  
[[-0.22893012  0.26759025 -0.8656726 ]  
 [-1.3631531  1.5353655  1.9220301 ]  
 [-1.766513  -1.1290098  1.993582  ]  
 [ 1.185827  -1.3770188 -0.81323206]  
 [ 1.1692326 -0.1399837  1.723129  ]]
```

K-clustering 결과

각각의 cluster의 중심인 mean 값입니다.

각 center들은 분포의 순서대로 출력된 것이 아니라
5개의 분포에서 뽑은 1500개의 sample로
clustering을 진행한 것이기 때문에
이후 각 분포가 어느 cluster를 따르는지
확인해주어야 합니다.

```
Select_cluster :  
[[114, 26, 28, 58, 74], [38, 1, 15, 142, 104], [4, 9, 125, 14, 148], [5, 120, 169, 1, 5], [6, 131, 31, 2, 130]]  
1 distribution == 1 cluster  
2 distribution == 4 cluster  
3 distribution == 5 cluster  
4 distribution == 3 cluster  
5 distribution == 2 cluster
```

각각의 분포에서 뽑은 300개의 sample이
어느 클러스터에 들어가는지 각 분포마다 체크한 것입니다.
이것을 보고 각각의 분포가 몇번째 cluster로 집중되는지 파악하였습니다.

02

Final

실행 결과

```
Select_cluster :  
[[114, 26, 28, 58, 74], [38, 1, 15, 142, 104], [4, 9, 125, 14, 148], [5, 120, 169, 1, 5], [6, 131, 31, 2, 130]]  
1 distribution == 1 cluster  
2 distribution == 4 cluster  
3 distribution == 5 cluster  
4 distribution == 3 cluster  
5 distribution == 2 cluster
```

예를 들면 첫번째 분포는 첫번째 cluster에 가장 많이 모이므로
첫번째 분포에 해당하는 cluster(class)는 첫번째 cluster이다.
라고 생각하고 이후 실험을 진행하였습니다.

```
Select_cluster :  
[[114, 26, 28, 58, 74], [38, 1, 15, 142, 104], [4, 9, 125, 14, 148], [5, 120, 169, 1, 5], [6, 131, 31, 2, 130]]  
1 distribution == 1 cluster  
2 distribution == 4 cluster  
3 distribution == 5 cluster  
4 distribution == 3 cluster  
5 distribution == 2 cluster
```

또한 두번째 분포의 점들을 보면
자신의 cluster 다음으로 sample들이 많이 분류된 cluster는
다섯번째 cluster인걸 확인할 수 있는데
이는 앞서 분포 그림에서 보았듯 두번째 분포와
세번째 분포가 겹치는 부분이 있고
그런 경향의 sample들이 많이 뿔었기 때문입니다..

02

Final

실행 결과

Max_dis & Min_dis :

```
[4.1716065267529405, 3.587479267342347, 3.522543641195532, 3.375766040679087, 4.182236413600973]  
[0.3366577542234803, 0.4993188279273909, 0.4373168940320196, 0.212810801692181, 0.26039314227993443]
```

이후 각각의 cluster로 분류된 점들 중
center와 가장 멀리 떨어져 있는 점을 이용해
해당 cluster의 반지름(max_distance) 으로 설정해주었습니다.
min_distance는 mean과 가장 가까운 data는 거리가 얼마나 되는지
측정하기 위해 구하였습니다.

```
First distribution's point (index : 302) is in original cluster
Second distribution's point (index : 302) is in original cluster
Third distribution's point (index : 302) is in original cluster
Forth distribution's point (index : 302) is in original cluster
Fifth distribution's point (index : 302) is in original cluster
```

```
First distribution's point (index : 303) is in original cluster
Second distribution's point (index : 303) is in original cluster
Third distribution's point (index : 303) is in original cluster
Forth distribution's point (index : 303) is in original cluster
Fifth distribution's point (index : 303) is in original cluster
```

```
Sixth distribution's point (index : 33) is in 1 cluster
Sixth distribution's point (index : 34) is in 1 cluster
Sixth distribution's point (index : 35) is in 1 cluster
Sixth distribution's point (index : 36) is in 1 cluster
Sixth distribution's point (index : 37) is in 4 cluster
Sixth distribution's point (index : 38) is in 1 cluster
Sixth distribution's point (index : 39) is in 4 cluster
Sixth distribution's point (index : 40) is in 2 cluster
```

Test 결과

기존의 5개의 분포에서 각각
clustering에 쓰이지 않은 100개의 sample을 추출하고

clustering에 사용하지 않은 1개의 분포에서
100개의 sample을 추출해

각각의 sample 들이 어느 cluster에 들어가는지 확인하였
습니다.

```
First distribution's point (index : 325) is in original cluster
Second distribution's point (index : 325) is in original cluster
Third distribution's point (index : 325) is in original cluster
Forth distribution's point (index : 325) is not in original cluster.(Instead in 2 cluster)
Fifth distribution's point (index : 325) is in original cluster
```

```
First distribution's point (index : 326) is in original cluster
Second distribution's point (index : 326) is not in original cluster.(Instead in 5 cluster)
Third distribution's point (index : 326) is in original cluster
Forth distribution's point (index : 326) is in original cluster
Fifth distribution's point (index : 326) is in original cluster
```

```
First distribution's point (index : 327) is in original cluster
Second distribution's point (index : 327) is in original cluster
Third distribution's point (index : 327) is in original cluster
Forth distribution's point (index : 327) is in original cluster
Fifth distribution's point (index : 327) is not in original cluster.(Instead in 5 cluster)
```

Test 결과

대체로 원래 분포가 갖는 cluster에 포함이 되었지만
간혹 다른 cluster로 분류되는 경우가 있었습니다.

이는 분포에서 generate된 sample들이
분포의 중심 주변에서만
위치해있지 않고

각각의 분포가 겹치는 부분 또한 있기 때문에
center까지의 거리가 원래의 분포보다
다른 분포의 center에 가까운 경우가 있기 때문입니다.

정확도 (잘못 분류될 확률)

1번 분포 – 0/100, 2번 분포 – 11/100,
3번 분포 – 13/100, 4번 분포 – 13/100
5번 분포 – 16/100

```
First distribution's point (index : 325) is in original cluster
Second distribution's point (index : 325) is in original cluster
Third distribution's point (index : 325) is in original cluster
Forth distribution's point (index : 325) is not in original cluster.(Instead in 2 cluster)
Fifth distribution's point (index : 325) is in original cluster
```

```
First distribution's point (index : 326) is in original cluster
Second distribution's point (index : 326) is not in original cluster.(Instead in 5 cluster)
Third distribution's point (index : 326) is in original cluster
Forth distribution's point (index : 326) is in original cluster
Fifth distribution's point (index : 326) is in original cluster
```

```
First distribution's point (index : 327) is in original cluster
Second distribution's point (index : 327) is in original cluster
Third distribution's point (index : 327) is in original cluster
Forth distribution's point (index : 327) is in original cluster
Fifth distribution's point (index : 327) is not in original cluster.(Instead in 5 cluster)
```

Test 결과

하지만 결과에서 볼 수 있듯이
각각의 분포와 겹치거나 가까운(주변에 있는) 분포에
해당하는 cluster로 분류되는 것을 볼 수 있습니다.

4번째 분포sample은 2번째 cluster(5번째 분포)
2번째 분포의 sample은 5번째 cluster(3번째 분포)
5번째 분포의 sample은 5번째 cluster(3번째 분포)

```
Sixth distribution's point (index : 33) is in 1 cluster  
Sixth distribution's point (index : 34) is in 1 cluster  
Sixth distribution's point (index : 35) is in 1 cluster  
Sixth distribution's point (index : 36) is in 1 cluster  
Sixth distribution's point (index : 37) is in 4 cluster  
Sixth distribution's point (index : 38) is in 1 cluster  
Sixth distribution's point (index : 39) is in 4 cluster  
Sixth distribution's point (index : 40) is in 2 cluster
```

Test 결과

cluster를 만드는데 사용되지 않은 6번째 분포의 sample은 대체로 가까운 분포 (1,2,5)와 겹치는 분포(1)의 cluster로

(1번째 분포 – 1번째 cluster,
2번째 분포 – 4번째 cluster
5번째 분포 – 2번째 cluster)

분류되는 경향이 있는 것을
확인 할 수 있었습니다.

```
Sixth distribution's point (index : 38) is in 1 cluster  
Sixth distribution's point (index : 39) is in 4 cluster  
Sixth distribution's point (index : 40) is in 2 cluster  
Sixth distribution's point (index : 41) is not in any cluster  
Sixth distribution's point (index : 42) is in 1 cluster  
Sixth distribution's point (index : 43) is in 2 cluster  
Sixth distribution's point (index : 44) is in 4 cluster
```

Test 결과

하지만

cluster를 만드는데 사용되지 않은 6번째 분포의 sample은 cluster의 center들과 너무 멀어서 어느 cluster로 분류되지 않는 모습도 볼 수 있습니다.

```
First distribution's point (index : 346) is in original cluster  
Second distribution's point (index : 346) is in original cluster  
Third distribution's point (index : 346) is not in original cluster.(Instead in 5 cluster)  
Forth distribution's point (index : 346) is in original cluster  
Fifth distribution's point (index : 346) is in original cluster
```

주의사항으로

조금 더 타이트하게 측정하기 위해

cluster의 max_distance보다 distance가 작을 경우에만

해당 cluster로 판단하였습니다.

이후 다른 cluster와의 비교를 할 때는

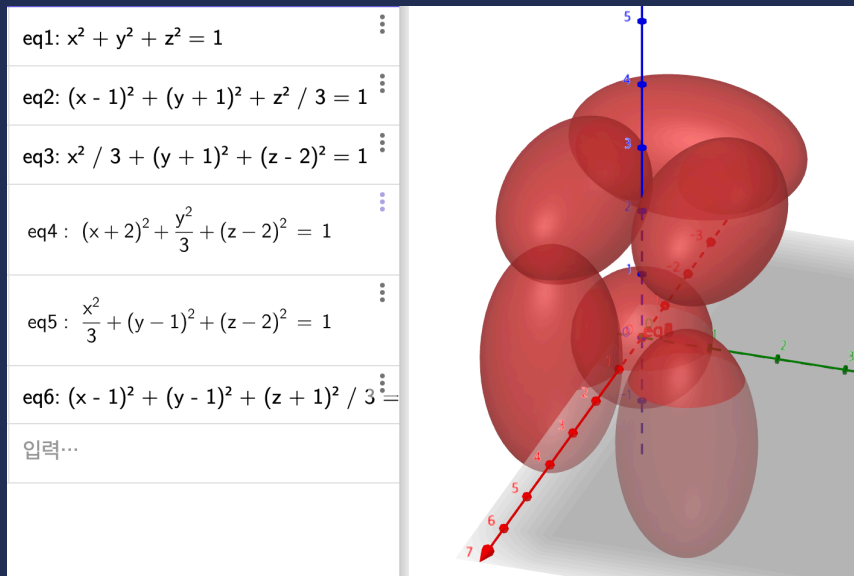
max_distance와 같은 경우에도 해당 cluster로 측정하였기 때문에

다음과 같이 처음에는 자신의 cluster로 분류되지 않았다가

나중에 cluster로 분류되는 경우가 있습니다.

03 Final

생각 해볼 점



First distribution's point (index : 302) is in original cluster
Second distribution's point (index : 302) is in original cluster
Third distribution's point (index : 302) is in original cluster
Forth distribution's point (index : 302) is in original cluster
Fifth distribution's point (index : 302) is in original cluster

First distribution's point (index : 303) is in original cluster
Second distribution's point (index : 303) is in original cluster
Third distribution's point (index : 303) is in original cluster
Forth distribution's point (index : 303) is in original cluster
Fifth distribution's point (index : 303) is in original cluster

느낀 점

실행의 결과로 해당 cluster들로 잘 분류되는 것으로 보아 sample의 수를 크게 해도 효과적일 것이라는 생각이 들었습니다.

mean과 variance를 조정해 분포의 거리를 멀게 하면 더욱 확실하게 classification이 가능할 테지만 overfitting 문제가 나타날 수도 있을 것 같습니다.

또한 분포의 중심에 몰려있는 sample들이 많이 뿔힌다면 확실한 classification이 가능 할 것 같습니다.

따라서 목적에 맞게 mean과 covariance를 조정하고 적절한 sample들을 뽑아 사용하는 것이 중요한 것 같습니다.