DBSCAN 알고리즘 개선하기

1. 실습 목표   
   과제 1에서 Eps, MinPts만 변경가능이라는 제약된 조건으로 DBSCAN Algorithm을 수행하였다.  
   마지막주차까지 배운 모든 내용을 사용해 과제1의 DBSCAN Algorithm의 V-measure을 개선하라.  
   (단, 올바르지 않을 경우 감점. → Ex. MinPts를 0으로 설정함과 같이 합리적이지 않을 경우)
2. 실습 내용

과제 1과 달리 변경가능 요소 제약 조건이 없기에, 추가적으로 PCA의 n\_component 값을 변경하여 V-measure을 높이도록 실습을 진행하였다. 따라서 고려한 변경 요소는 n\_component, Eps, MinPts로 총 세 가지였는데, 이때 학습 데이터셋의 크기가 그리 크지 않음과 차원의 저주를 고려해 n\_component를 10차원 이하로 설정하였다. 해당 변경 범위에서의 결과는 아래와 같았다.

|  |  |  |
| --- | --- | --- |
| 입력값 | | 출력값 |
| n\_component | Eps | V-measure |
| 2 | 0.0136 | 0.7193051428901487 |
| 3 | 0.0534 | 0.8133647393409461 |
| 4 | 0.0784 | 0.8290678795188465 |
| 5 | 0.1215 | 0.8485325103015419 |
| 6 | 0.1190 | 0.8430434435904525 |
| 7 | 0.2854 | 0.8399905844570222 |
| 8 | 0.3071 | 0.8308202957590556 |
| 9 | 0.3108 | 0.8299961673402556 |
| 10 | 0.2547 | 0.8242888731211486 |

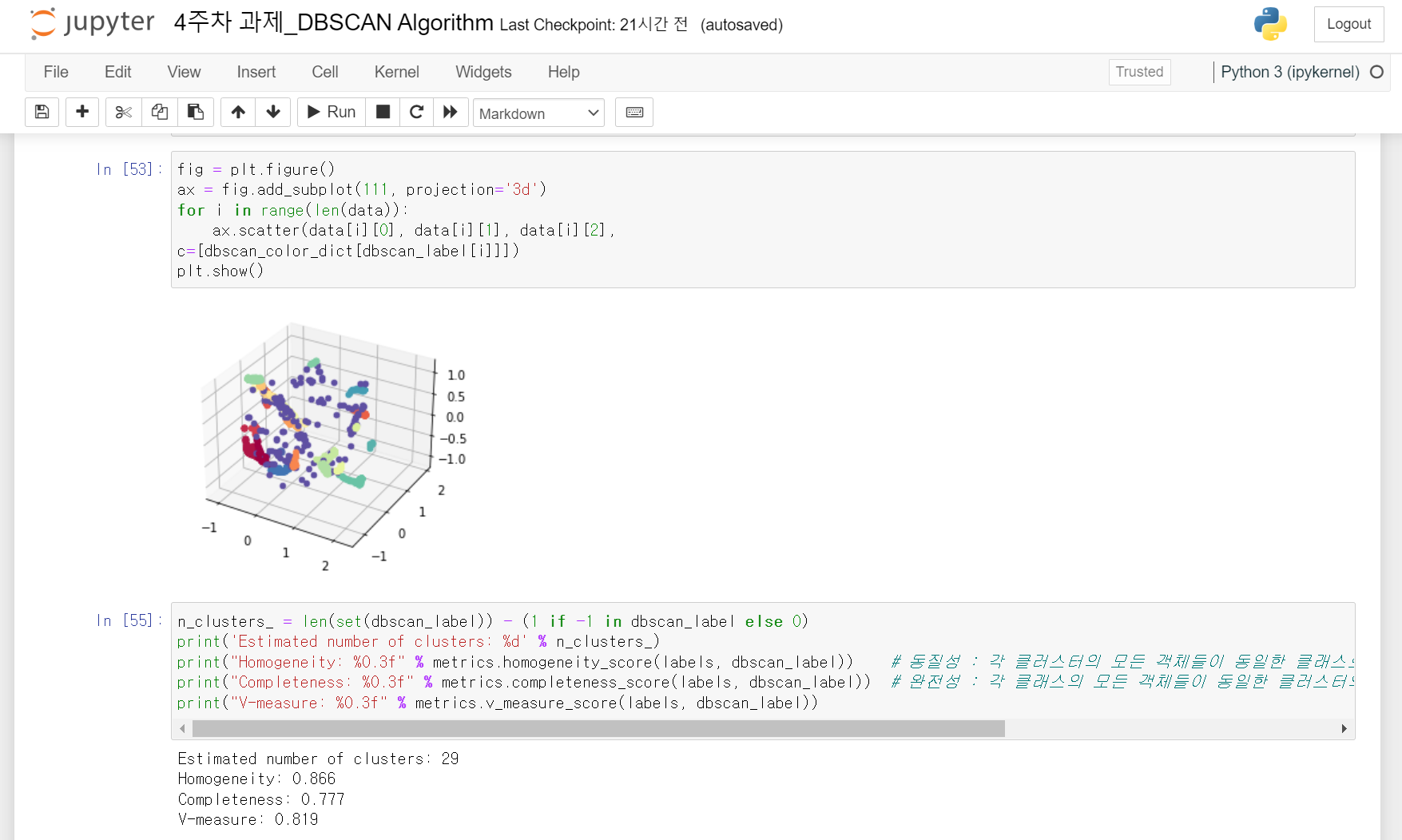
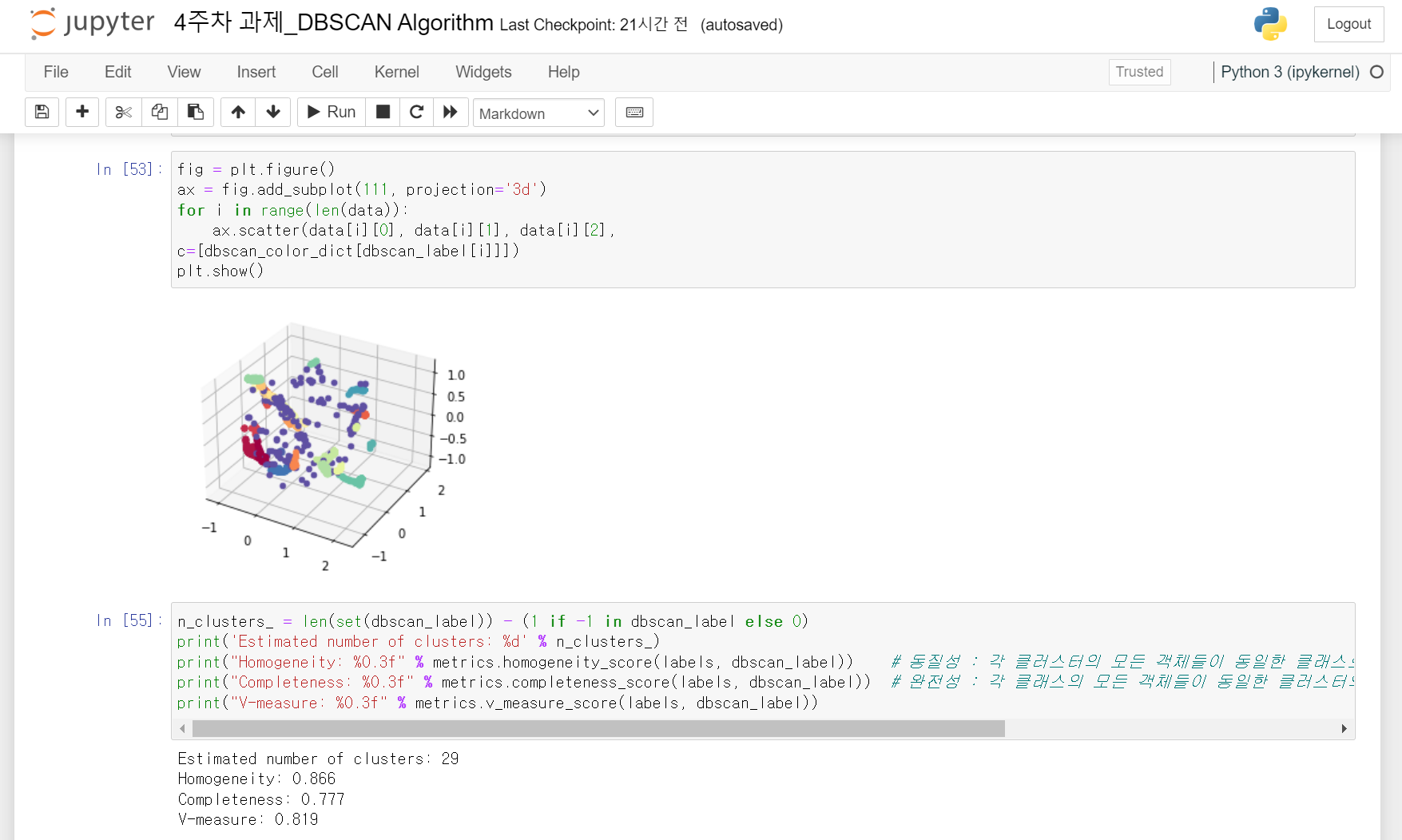
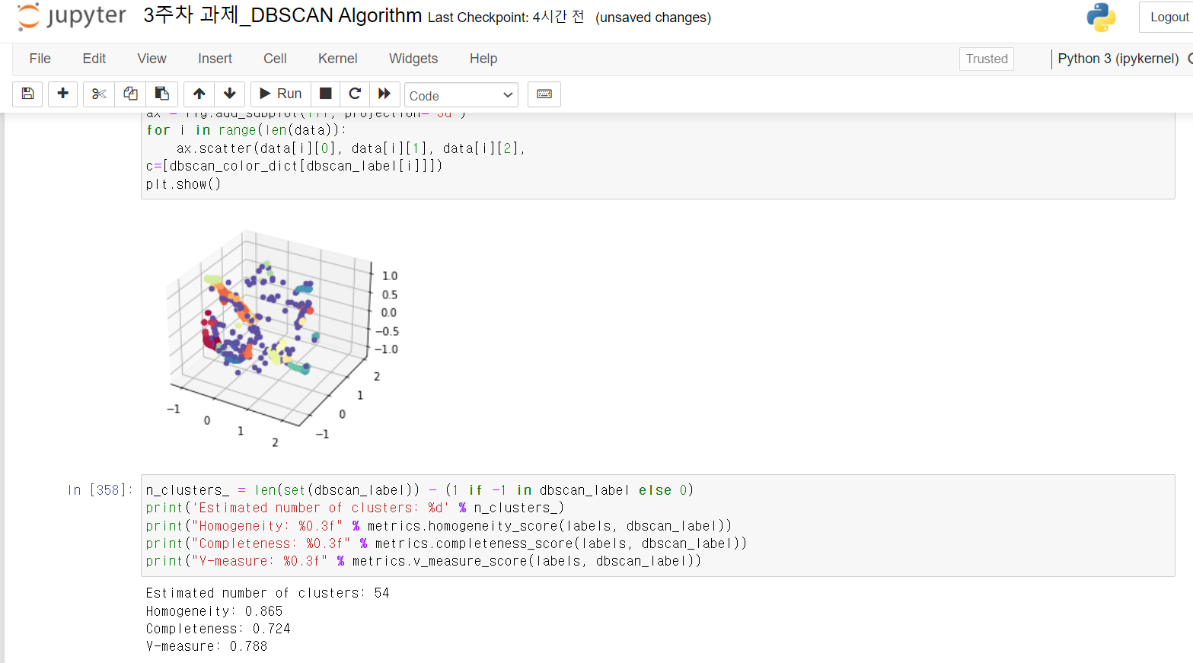
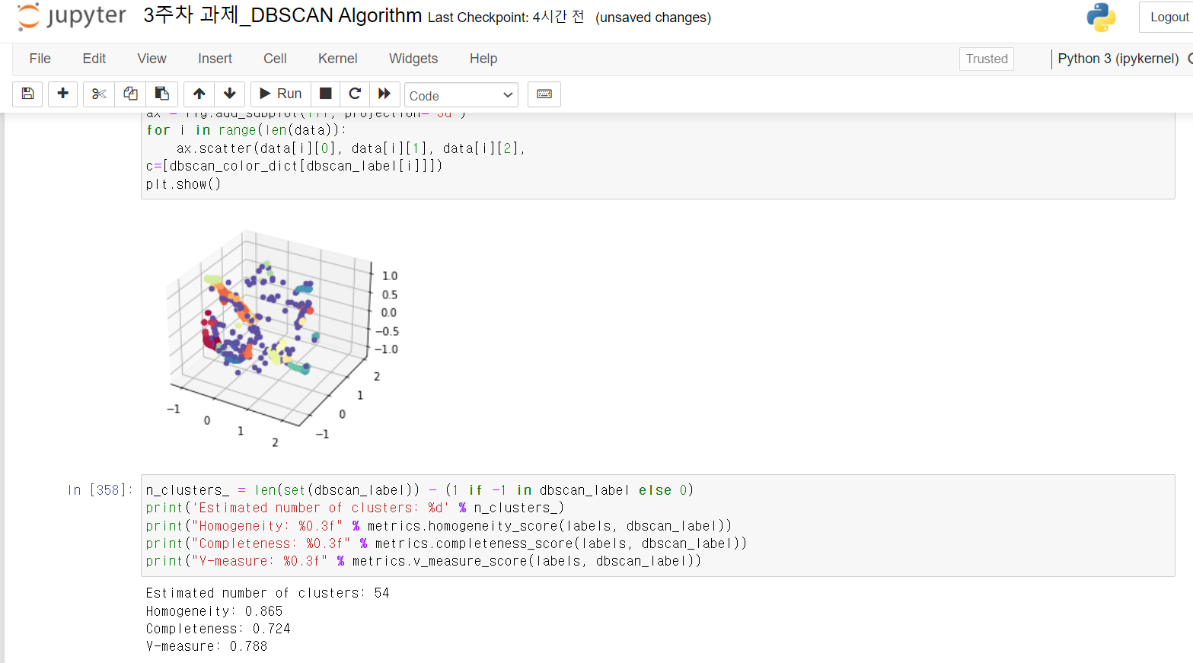
[표1] MinPts=1일 경우, V-measure의 최댓값  
  
임의로 MinPts를 1로 고정하고, n\_component와 Eps를 변경하며 찾은 V-measure의 최댓값은 [표1]과 같았다. 도출된 n\_component와 V-measure 값의 관계성을 통해 n\_component=5일 때를 기준으로 양 쪽으로 점점 V-measure의 값이 감소함을 알 수 있었고 때문에 n\_component=5가 경사상승법의 극점임을 알 수 있었다.  
이는 과제 1의 V-measure 결과값보다 높은 수치의 V-measure을 가지긴 하나, MinPts가 1일 경우 모든 포인트가 코어 포인트가 되기에 합리적 설정값이라 보기 어렵다. 따라서 MinPts가 1인 고정 범위가 아닌, Rule of Thumb에 따라 범위를 n\_component+1이상 2\*n\_component이하로 재설정 후 최대 V-measure을 탐색해주었다. 해당 결과는 아래와 같았다.

|  |  |  |  |
| --- | --- | --- | --- |
| 입력값 | | | 출력값 |
| n\_component | MinPts | Eps | V-measure |
| 2 | 3 | 0.0136 | 0.7193051428901487 |
| 3 | 4 | 0.0534 | 0.8133647393409461 |
| 4 | 5 | 0.0784 | 0.8290678795188465 |
| 5 | 6 | 0.1215 | 0.8485325103015419 |

[표 2] n\_component+1<= MinPts <=2\*n\_component 일 경우, V-measure의 최댓값

MinPts를 1로 고정 후 진행했던 [표1]과 달리, 샘플 수를 늘리자 모든 V-measure의 값이 각각소폭 감소한 것을 볼 수 있었다. 그럼에도 V-measure의 값 분포는 유사하게 나타남을 확인할 수 있었다. 따라서 여전히 n\_component=5일 때 극값, 즉 최대 V-measure값을 가지는데, 이때 MinPts의 경우 6으로 합리적 설정값 범위 내에 존재하므로 해당 V-measure값은 적합하게 개선된 결과값이라 할 수 있다.

1. 실습 결과  
   ▪ 성능 개선 전 실행 결과   
   → n\_component=3, MinPts=3, Eps=0.1215, V-measure=0.788  
     
   ▪ 성능 개선 후 실행 결과  
     
   → n\_component=5, MinPts=6, Eps=0.049, V-measure=0.819



1. 결론 및 향후 실습 방향  
   적합한 n\_component와 MinPts, Eps 값의 설정을 통해, V-measure 값을 높이고 DBSCAN 알고리즘의 성능을 개선할 수 있었다. 허나 개발환경 및 시간적 제한 사항, 능력 부족의 원인으로 인해 차원을 10차원 이하로 제한하거나 Eps를 소수점 네 자리로 제한하는 등 모든 경우의 수를 고려하지 않은 채 탐색을 진행하였다.  
   따라서 구한 값이 개선된 설정 값임은 맞으나, Global maximum값이 아닌 Local maximum일 수 있다는 문제점이 있다. 보다 뛰어난 성능 개선을 위해선 이를 고려하여 설정 값들을 탐색해야 할 것이다.