

[머신러닝특론 HW3 - 가우시안혼합모델, 군집화]

- 제출기한
- 5월 01일 09:00 ~ 5월 12일 23:59 까지
- 연장제출기한
- 5월 13일 00:00 ~ 5월 19일 23:59 까지 (연장제출은 20% 감점되어 반영됨)
- 과제수행 및 제출방법
- Python 혹은 Matlab/Octave를 사용하여 과제 수행.
- 코드파일과 수행결과보고서를 모두 제출하셔야 합니다.
(Colab을 사용하시는 경우, ipynb 파일 하나로 제출하셔도 됩니다.)
- 코드파일 혹은 보고서 둘 중 하나에는 코드에 대한 상세한 주석(설명)이 있어야 합니다.
- 코드파일(.m 혹은 .py 파일)과 보고서(word or 한글 or pdf)를 zip압축하여 제출
>> (예) HW3_학번_성명.zip (ZIP파일로 압축하여 제출)

1. (50점) 참고교재의 프로그램 7-1(강의 자료 Matlab/Octave 코드)을 참고하여, K-means 군집화 알고리즘을 수행해 보시오. (참고교재 프로그램 7-1 코드 오타 관련 수정사항은 자료실의 [참고교재PDF 오타관련] 파일을 참고하시기 바랍니다.)

(1) (5점) (데이터 생성) 다음과 같은 평균과 공분산을 가지는 2차원 데이터를 클래스 당 100개씩 생성하고, 이를 2차원 평면에 표시하시오(총 3개의 클래스를 생성하고, 클래스별로 색을 다르게 해서 plot할 것).

$$\mu_1 = \begin{pmatrix} 0 \\ 4 \end{pmatrix}, \mu_2 = \begin{pmatrix} 4 \\ 4 \end{pmatrix}, \mu_3 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

(2) (10점) (1)에서 생성한 데이터의 각 클래스의 평균을 계산하고, 이를 (1)의 평면에 함께 표시하시오.

(3) (35점) (1)에서 생성한 데이터에 K-means 알고리즘(K=3으로 설정)을 적용하여 군집화를 수행하시오. 이때 알고리즘 반복횟수는 10으로 한정하고, 각 반복에 따른 군집 결과 및 군집의 중심을 plot을 활용하여 출력하시오. 그리고 마지막 수행 결과 출력에서는 (2)에서 계산하였던 평균값도 함께 출력하여 군집화를 통해 획득한 군집의 중심과 실제 데이터가 가진 중심 간의 차이가 어떻게 나타나는지를 관찰하시오. (최종적으로 최적의 군집화 상태에 도달하기 까지 알고리즘이 몇 번 반복되었는지 언급하고, 각 반복 step마다 plot 해야 함)

2. (50점) 가우시안 혼합 모델을 사용하여 주어진 데이터 HW3Data.mat에 대한 분석을 수행하시오.

(1) (10점) 데이터 HW3Data.mat을 불러온 다음 2차원 평면에 산점도를 그리시오.
(scatter 함수 사용할 것)

(2) (40점) 참고교재의 프로그램 10-1을 참고하여, 주어진 데이터에 대한 가우시안 혼합 모델을 적용하여 분석을 수행하시오. 이때 가우시안 성분의 수는 {2, 6, 10} 으로 변경해 가며 수행하고, 각각의 최종 결과를 산점도로 출력하여 이를 비교한 후 그 결과를 서술하시오.

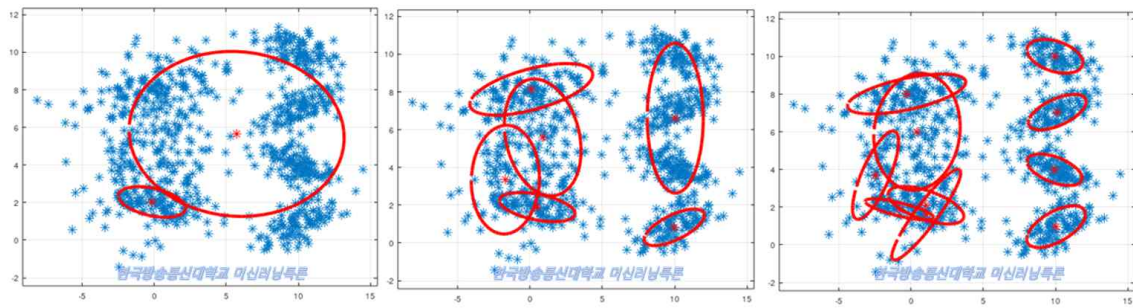


그림1. 가우시안 성분 수에 따른 분석 결과 예시
(가우시안 성분의 수는 왼쪽부터 2, 6, 10)