

## [머신러닝특론 HW2 - Bayes Classifier & K-NN Classifier]

- 제출기한

- 4월 04일 12:00 ~ 4월 14일 23:59 까지

- 연장제출기한

- 4월 15일 00:00 ~ 4월 21일 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) HW2\_학번\_성명.zip (ZIP파일로 압축하여 제출)

### [문제1 (총 40점)]

참고교재의 프로그램 4-1을 참고하여 다음 주어진 단계에 따라 데이터를 생성하고 베이지안 분류기를 이용하여 분류를 수행하시오.

- (1) (20점) 다음과 같은 평균과 공분산을 가지는 가우시안 분포를 따르는 각 100개씩의 2차원 데이터로 이루어진 두 클래스 집합  $C_1$ ,  $C_2$ 를 생성하고, 이를 2차원 평면상에 산점도를 그리시오. (그림1 참고, 데이터의 산점도 형태는 random의 영향으로 다를 수 있음)

$$\mu_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}, \mu_2 = \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix}$$

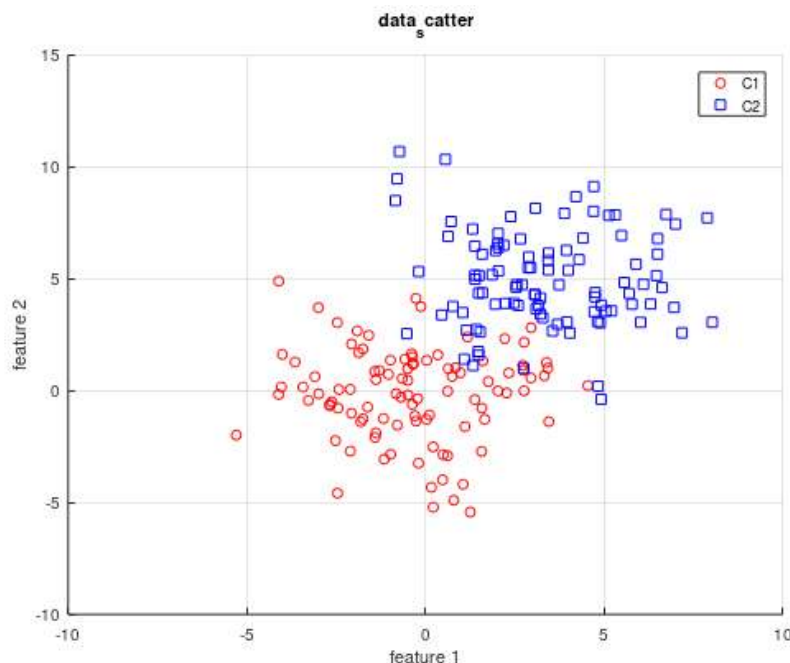


그림 1. 생성된 데이터의 산점도.

- (2) (20점) 베이지 분류기에서 클래스 공통 공분산행렬을 가정한 경우와 일반적인 공분산행렬을 가정한 경우의 판별함수를 프로그래밍으로 구현하고 이를 이용하여 각각의 경우 두 개의 새로운 데이터  $x_1 = [1, 2]^T$  와  $x_2 = [0, -2]^T$ 가 어떤 클래스에 속하는지를 판단하시오. 또한, 문제 1-(1)에서 출력한 산점도에 두 새로운 데이터를 각자 다른 모양과 색으로 구분되도록 함께 plot하시오.

**[문제2 (총 60점)]**

참고교재의 프로그램 5-1을 참고하여 K-NN 분류기로 iris 데이터를 분류하시오.

- (1) (30점) 각 클래스당 데이터 50개씩 총 3개의 클래스로 이루어진 iris 데이터를 LMS 과제 게시판에서 다운로드 받고, K-근접이웃 분류기를 이용하여 iris 데이터를 분류하시오. 데이터의 거리 계산은 유클리디안 거리(euclidean distance)를 사용하고, K값을 5, 10, 15, 20, 25, 30으로 변화시키면서 수행하여 K값에 따른 학습데이터의 오분류율(misclassification rate)을 표로 정리 및 비교하시오.

- (2) (30점) 주어진 iris 데이터 150개를 각각 학습데이터 120개, 테스트데이터 30개로 나누시오. (나누는 방법은 자유롭게 할 것) 거리함수와 K값의 변화를 2-(1)과 동일하게 설정하여 테스트 데이터에 대한 오분류율을 표로 정리하고 이의 변화 양상을 2-(1)의 결과와 비교하시오.