

[머신러닝특론 HW1 - Feature Extraction]

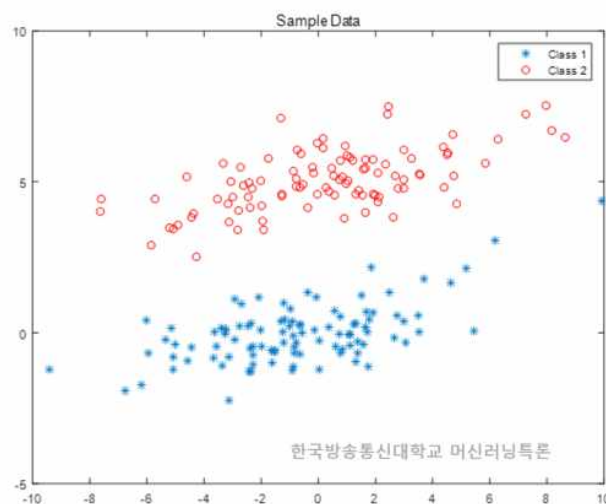
- 제출기한
 - 3월 22일 00:00 ~ 4월 01일 23:59 까지
- 연장제출기한
 - 4월 02일 00:00 ~ 4월 08일 23:59 까지 (연장제출은 20% 감점되어 반영됨)
- 과제수행 및 제출방법
 - Python 혹은 Matlab/Octave을 사용하여 과제 수행.
 - 코드파일과 수행결과보고서를 모두 제출하셔야 합니다.
 - 코드파일 혹은 보고서 둘 중 하나에는 코드에 대한 상세한 주석(설명)이 있어야 합니다.
 - 코드파일(.m 혹은 .py 파일)과 보고서(word or 한글 or pdf)를 zip압축하여 제출
>> (예) HW1_학번_성명.zip (ZIP파일로 압축하여 제출)
- 유의사항
 - 연장제출기한이 지나면 과제를 제출하셔도 받지 않습니다.
 - 타 원우의 과제를 copy&paste한 경우에는 0점처리 됩니다.
 - Python을 사용하실 경우, sklearn과 같은 간편한 라이브러리에 이미 구현된 PCA등을 사용 하시면 안되며, 반드시 자신이 직접 PCA와 LDA를 구현하셔야 합니다.

[문제1 (총 40점)].

참고교재의 프로그램 8-1(강의자료 MATLAB 코드)을 참고하여, 특징 추출을 수행해 보시오.
(단, 산점도 출력시 `axis([-10 10 -5 10])` 명령을 사용하여 출력 공간을 통일하여 표시.)

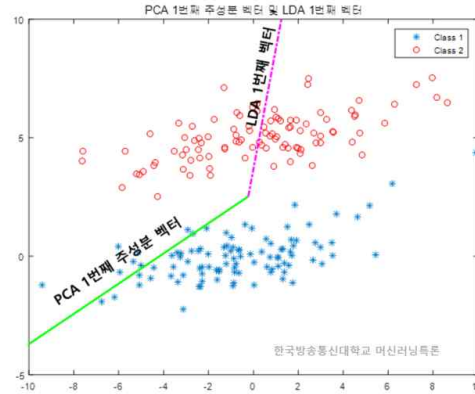
- (1) (10점) 다음과 같은 분포를 가진 두 클래스의 데이터를 생성하고 아래 그림과 같이 산점도를 그리시오. 이때 샘플의 개수는 각 클래스에 대해 100개씩 총 200개입니다.

$$\mu_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mu_2 = \begin{pmatrix} 0 \\ 5 \end{pmatrix}, \Sigma_1 = \Sigma_2 = \begin{pmatrix} 10 & 2 \\ 2 & 1 \end{pmatrix}$$



<그림 1. 특징 추출을 위한 샘플 데이터 생성 예>

- (1) (20점) (1)에서 생성한 데이터에 대해 PCA를 적용하여 첫 번째 주성분 벡터를 찾고 이를 아래 그림2와 같이 2차원 평면상에 표시하시오. 이를 LDA에 대해서도 수행하고 함께 출력하시오.



<그림 2. PCA 및 LDA 벡터>

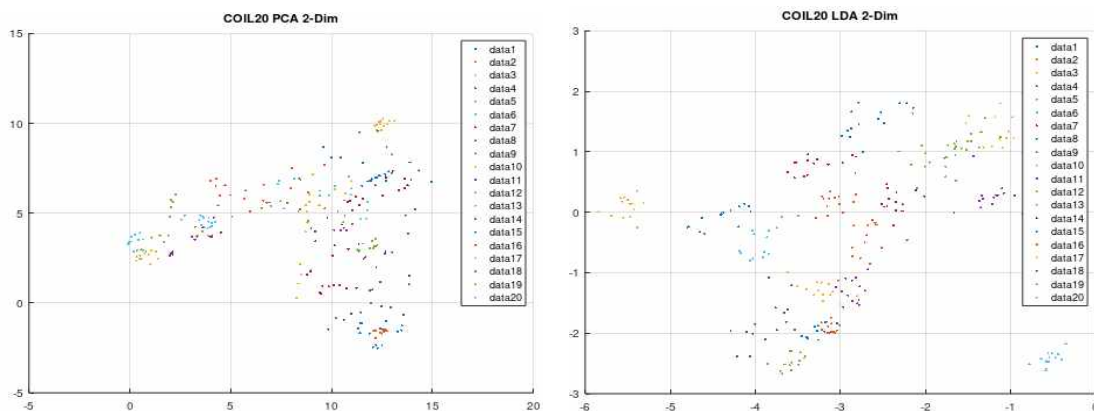
- (1) (10점) PCA의 결과와 LDA의 결과를 그림으로 비교하고 두 결과의 차이를 간단히 서술하시오.

[문제2 (총 60점)]

첨부된 COIL20 데이터(Columbia Object Image Library, HW1_COIL20.mat)에 대한 특징 추출 및 2차원 출력을 수행하기 위한 시스템을 구성하시오. (첨부된 HW1_COIL20.mat에는 학습 및 테스트 데이터 모두 포함되어 있으므로, 이중 **학습 데이터에 해당하는 데이터만 사용할 것**. X는 학습 데이터, Y는 학습 데이터의 클래스 정보에 해당함.)

(※ COIL20 데이터에 대한 자세한 사항은 첨부된 자료 'COIL20.pdf'를 참고 하세요.)

참고교재의 프로그램 9-1을 참고하여 COIL20 데이터를 PCA, LDA를 통한 2차원 특징추출하고 이를 각각 2차원 공간에 출력하시오. 즉, 먼저 PCA를 통해 2차원 특징 추출하여 2차원 공간에 출력하고, 다음으로는 LDA를 통해 2차원으로 특징추출 하되, 정보보존을 95%로 PCA를 우선 수행하여 획득한 특징값에 대해 LDA를 수행하여 2차원 LDA 특징값을 추출하여 2차원 공간에 출력할 것. 2차원 특징값 출력 시 아래 그림과 같이 각 클래스가 구별될 수 있도록 출력하시오. (클래스별 출력은 예제 코드 참고, exam_plotmulticlass.m)



<그림 3. COIL20 데이터 PCA 및 LDA 2차원 특징 추출 후 출력>