

머신러닝특론 과제1 리포트

서론

본 보고서는 특징 추출 기법인 주성분 분석(PCA)과 선형 판별 분석(LDA)을 다양한 데이터셋에 적용하고 그 결과를 분석합니다. 문제 1에서는 주어진 통계적 분포를 갖는 두 클래스 데이터를 생성하여 PCA와 LDA를 적용하고, 그 결과를 시각적으로 비교합니다. 문제 2에서는 실제 이미지 데이터셋인 COIL20 데이터에 대해 PCA와 LDA를 활용하여 2차원으로 특징을 추출하고, 각 클래스를 구별하여 시각화합니다.

문제 1 PCA, LDA 구현 및 차원 축소 연습

문제 1.1 산점도 그리기

본 문제에서는 다음과 같은 평균과 공분산 행렬을 갖는 두 개의 클래스 데이터를 각각 100 개씩 생성하였습니다.

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

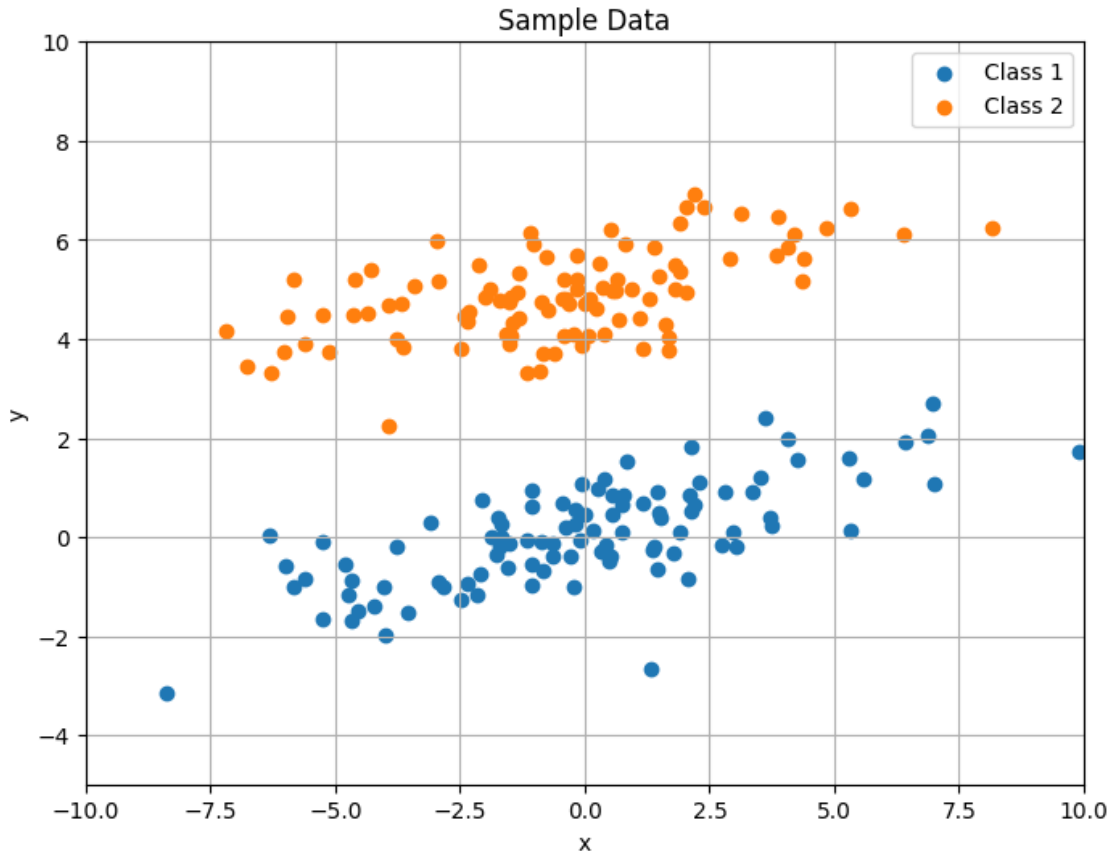
생성된 데이터의 산점도는 <그림 1>과 같습니다. 각 클래스는 서로 다른 색상과 마커로 표시되었으며, `axis([-10 10 -5 10])` 명령을 사용하여 출력 공간을 통일하였습니다.

무작위 데이터를 생성하기 위하여 Gaussian 분산 형태로 랜덤 값을 생성하는 numpy의 `np.random.multivariate_normal()` 함수를 사용하였습니다.

Source code : `pca_lda1_1.py`

결과

- 클래스 1 : (0,0)을 중심으로 퍼져있는 형태
- 클래스 2 : (0,5)를 중심으로 퍼져있는 형태



<그림 1>

문제 1.2 PCA 및 LDA 적용 및 시각화

앞서 문제 1.1 에서 생성한 데이터에 대해 PCA 와 LDA 을 진행하여 첫번째 주성분 벡터를 구하고, 그 결과를 그래프 상에 표시하는 작업을 진행 하였습니다.

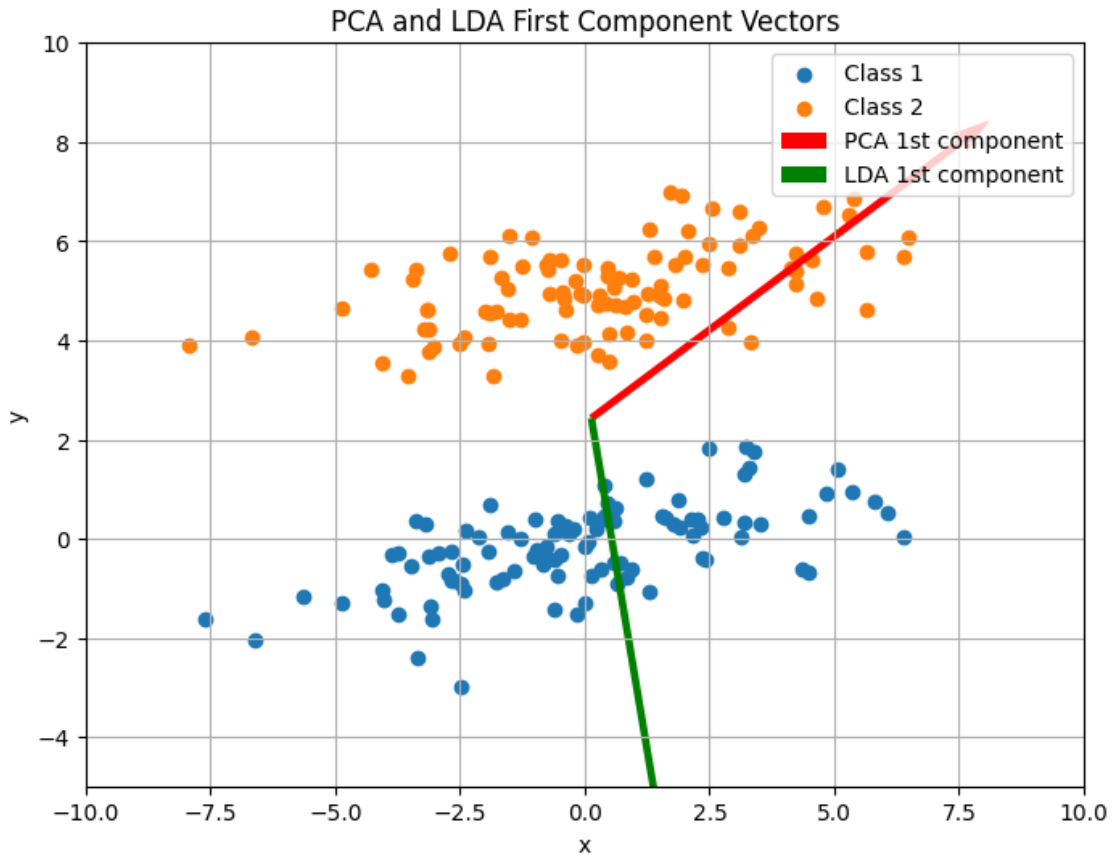
PCA, LDA 를 계산하기 위해 필요한 공분산은 numpy 의 `np.cov()` 함수를 사용하여 구하였습니다.

또한 고유값과 고유벡터를 계산하기 위해서 numpy 의 `np.linalg.eig()` 함수를 사용하였습니다. 해당 함수는 공분산행렬 A가 주어졌을 때 아래의 방정식을 풀어어내는 함수입니다. 그 결과로서 고유벡터 v 와 고유치 λ 가 계산 됩니다.

$$A * v = \lambda * v$$

Source code : `pca_lda1_2.py`

결과



<그림 2>

문제 1.3 PCA 와 LDA 의 비교

<그림 2>에서 볼 수 있듯이, PCA의 주성분 벡터는 데이터의 특징을 잘 보존할 수 있는 방향 (분산이 큰 방향)을 향하고 있는 반면에, LDA의 주성분 벡터는 데이터를 차이를 잘 구분할 수 있는 방향으로 향하고 있는 것을 알 수 있습니다.

문제 2: COIL20 데이터에 대한 특징 추출 및 2차원 출력 시스템 구축

COIL20 데이터셋은 20개의 서로 다른 물체를 다양한 각도에서 촬영한 이미지들로 구성되어 있습니다. 각 물체는 72개의 다른 시점에서 촬영되었으며, 본 과제에서는 학습 데이터에 해당하는 부분만을 사용하여 특징 추출을 수행합니다.

첨부된 HW1_COIL20.mat 파일에서 학습 데이터(X), 클래스정보(Y)를 불러와 PCA를 적용하여 2차원으로 축소하였습니다.

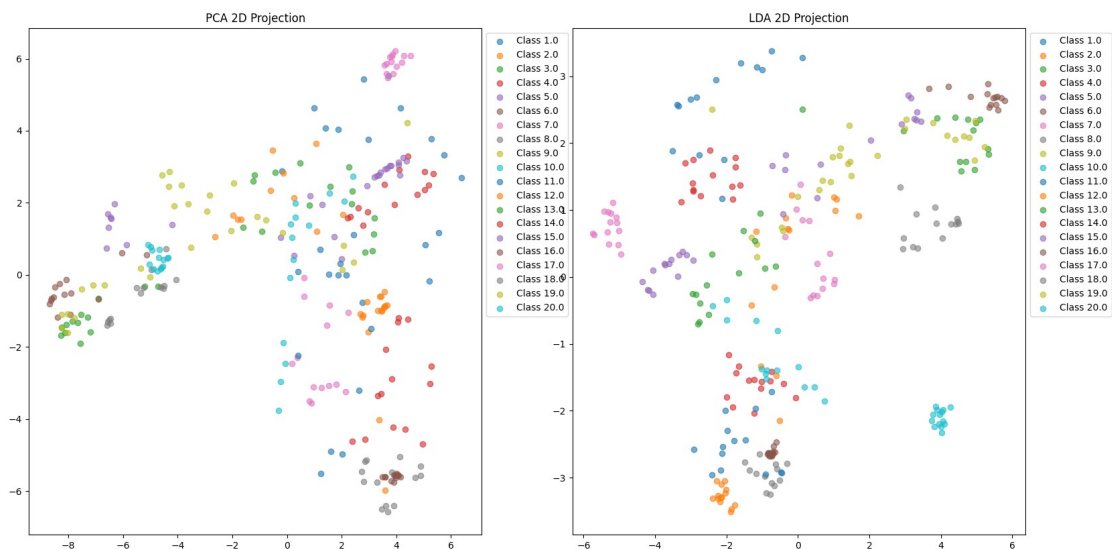
앞선 문제1과 다른점은 COIL20 데이터에는 20개의 클래스가 존재 하기 때문에 앞서 만들었던 함수를 확장하여 2개 이상의 클래스에도 대응하도록 하는 것입니다.

정보 보존율 95%를 계산하는 코드는 아래와 같은데, np.cumsum() 함수로 고유치의 누적합의 배열을 만들고, 그것을 전체 고유치의 합으로 나누어 고유치 누적합의 성분이 전체에서 차지하는 비율을 조사하여 95% 경계를 찾아냅니다.

```
cumulative_variance = np.cumsum(eigenvalues) / np.sum(eigenvalues)
```

Source code : pca_lda2.py

결과



<그림 3>