고급소프트웨어실습I

3주차 보고서

20171646 박태윤

1. 측정 결과

- 2차원

텍스트이(가) 표시된 사진

자동 생성된 설명

- 3차원

텍스트이(가) 표시된 사진

자동 생성된 설명

- 4차원

텍스트이(가) 표시된 사진

자동 생성된 설명

- 32차원

텍스트이(가) 표시된 사진

자동 생성된 설명

2. 분석

코드는 다음과 같다.

텍스트이(가) 표시된 사진

자동 생성된 설명

표시된 값에 따라 몇 차원으로 축소한 결과를 복원할지를 정할 수 있다.

student\_pca를 이용하여 기존의 데이터를 원하는 차원으로 pca할 수 있다. 기존의 데이터는 총 64차원의 데이터인데, 이를 2차원으로 줄인다고 하면 2개의 고유벡터와의 내적을 통해 2개의 축으로 정사영을 시켜서 데이터 분포의 차원을 줄이는 것이고, 32차원으로 줄인다고 하면 그보다 더 많은 32개의 축으로 데이터를 정사영 시키는 것을 의미한다. pca는 값의 차원을 낮추고 가장 중요한 정보들을 보존하는 식으로 구현을 하는데 이는 곧 고유값이 높은 순으로, 분산이 큰 순대로 고유벡터를 정렬하여 이 중 높은 순대로 고유벡터를 선택하는 것을 통해 차원을 변화를 시킨다. 결국 2차원으로 줄이는 것이 32차원으로 줄이는 것에 비해 훨씬 더 적은 수의 벡터로 현재 64차원의 데이터를 정사영 시켜서 표현하는 것이기에 기존 데이터와 차이가 커질 수밖에 없으므로 32차원으로 줄인 이후에 64차원으로 복원을 한 데이터가 2차원으로 줄여 64차원으로 복원을 한 데이터보다 오차율이 적다. 즉, 벡터의 수가 정보의 보존율을 나타낸다고 볼 수 있기에 계산 과정에서 고르는 고유 벡터의 수와 비례하는 차원 축소의 정도에 따라 오차율이 달라질 수 밖에 없다.