고급소프트웨어실습1

Lecture 3 과제

PCA reconstruction

분반 : 4

학번 : 20171669

이름 : 이재영

1. PCA reconstruction (image)

* 2차원으로 축소한 결과 복원 결과

텍스트이(가) 표시된 사진

자동 생성된 설명

* 3차원으로 축소한 결과 복원 결과

텍스트, 낱말맞추기게임이(가) 표시된 사진

자동 생성된 설명

* 4차원으로 축소한 결과 복원 결과

텍스트이(가) 표시된 사진

자동 생성된 설명

* 32차원으로 축소한 결과 복원 결과

텍스트, 낱말맞추기게임이(가) 표시된 사진

자동 생성된 설명

1. PCA reconstruct – code

* 2차원 축소 결과 복원 code

텍스트이(가) 표시된 사진

자동 생성된 설명

* 3차원 축소 결과 복원 code

텍스트이(가) 표시된 사진

자동 생성된 설명

* 4차원 축소 결과 복원 code

텍스트이(가) 표시된 사진

자동 생성된 설명

* 32차원 축소 결과 복원 code

텍스트이(가) 표시된 사진

자동 생성된 설명

1. 분석결과

PCA는 주성분 분석(Principal Component Analysis)라 불리며 차원 축소 알고리즘이다. PCA는 데이터에 가장 가까운 평면을 구한 뒤 데이터를 투영시키는 방법을 가진다. 이를 통해 데이터를 차원에 맞게 축소할 수 있고 이번 과제에서는 차원에 따라 축소한 결과들을 다시 원래 데이터로 복원했을 때의 비교를 주된 목적으로 한다. 이 때, 각각의 복원데이터와 기존 데이터 간의 MSE(mean square error), 즉 평균제곱오차를 비교한다. 먼저 MSE란 실제 데이터의 값과 추정값과의 차이(오차)의 제곱에 대한 평균을 취한 값으로 MSE가 작을수록 그 추정에 대한 정확성이 높아진다. 위 실험에 대한 결과를 보면 2차원 축소를 복원했을 때의 MSE는 858.94478…, 3차원의 경우 717.235244…, 4차원의 경우 616.191130…, 32차원의 경우 40.42470… 의 값을 확인할 수 있다. 또한 오차율에 관해서 PCA의 경우 원본 데이터에 대해 분산을 가장 많이 보존하는 축부터 찾으면서 투영할 차원만큼의 축을 선택을 한다. 투영시킬 때, 다차원일수록 분산을 보존하는 축을 많이 선택하므로 오차율이 더 적어지는 것을 볼 수 있다. 기존 디지털 데이터가 64차원인 것을 고려해보면 64차원에 가까운 차원일수록 MSE 값이 작아진다는 것을 확인할 수 있다. 즉, 큰 차원으로 축소한 후 복원했을 때, 기존 데이터와 비슷하다고 볼 수 있다. 복원된 사진결과를 통해서도 차원이 감소했다가 다시 복원되는 과정에서 데이터의 손실이 발생해서 기존과 조금의 차이가 나는데 큰 차원으로 축소되었다가 복원될 때가 가장 데이터의 손실이 적어 기존 데이터의 이미지와 유사하다는 것을 볼 수 있다.