**고급소프트웨어 실습 7주차 과제**

20171666 이예은

Eigenvector의 개수는 원래 데이터의 차원(dimension)과 같으므로 모든 eigenvector를 이용하여 원래 데이터를 투영하면 데이터 압축의 효과를 볼 수 없다. 또한 데이터의 분산 특성을 고려하지 않고 특정 몇 개의 eigenvector를 이용하여 투영시키면 데이터의 유용한 정보가 손실될 수 있다. 유용한 정보를 손실하지 않으면서 최대로 압축 효과를 얻을 수 있도록 eigenvector의 개수를 설정할 수 있는 방법이 있는지 설명해 보자.

* 데이터의 특성을 알기 위해서는 eigenvector와 eigenvalue가 필요하다. Eigenvalue의 값이 크면, 해당 eigenvalue로부터 생성된 eigenvector가 생성하는 기처 축에 데이터의 패턴이나 성질 등이 많이 표현되어 있다고 간주할 수 있다. 즉, 가장 큰 eigenvalue로부터 생성되는 eigenvector가 데이터에 대해 가장 많은 정보를 갖고 있다. 이를 이용해 해당 eigenvector가 아닌 부분들은 데이터 표현에 큰 영향을 주지 않는 부분들이므로 제거하거나, 무시함으로써 데이터를 압축할 수 있다. 그러므로 가장 큰 eigenvalue의 eigenvector만 남도록 차원 및 개수를 설정하면 정보를 손실하지 않으면서 최대로 압축 효과를 얻을 수 있다.