**고소실 7주차 과제**

**20171672 이정원**

**Q. Eigenvector의 개수는 원래 데이터의 차원(dimension)과 같으므로 모든 eigenvector를 이용**

**하여 원래 데이터를 투영하면 데이터 압축의 효과를 볼 수 없다. 또한 데이터의 분산 특성**

**을 고려하지 않고 특정 몇 개의 eigenvector를 이용하여 투영시키면 데이터의 유용한 정보**

**가 손실될 수 있다. 유용한 정보를 손실하지 않으면서 최대로 압축 효과를 얻을 수 있도록**

**eigenvector의 개수를 설정할 수 있는 방법이 있는 지 설명해 보자.**

데이터의 유용한 정보를 손실하지 않기 위해 가장 설명력이 높은 eigenvector을 선정해야 한다. 데이터에 대해 N\*N covariance matrix을 계산한 뒤 eigenvalue decomposition(A\*B = Bλ에서 B의 역행렬을 양변에 곱하는 과정)을 수행한다. 이 과정에서 covariance matrix을 가장 잘 설명하는 벡터를 찾는데, 보통 eigenvalue의 weight가 전체 weight 중 90%가 넘는(weight의 평균치를 내어 특정 기준을 선정하는 방법도 있다.) vector을 선택하는 방법이 있다. 이후 해당 space로 정사영 시켜 압축 효과를 낼 수 있을 것이다.