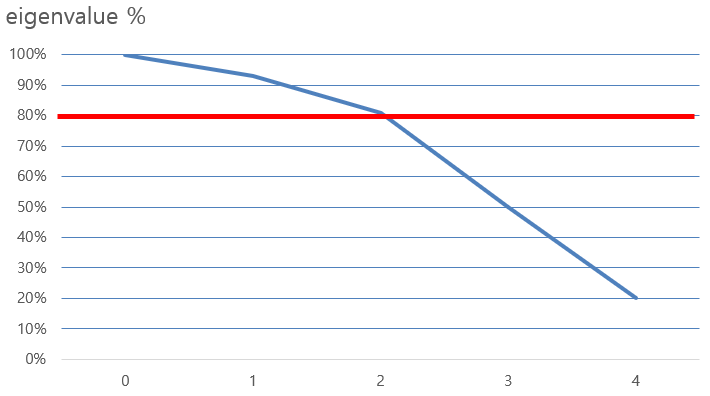
[고소실\_7주차과제]2반\_20161595\_배성현

1**. Eigenvector의 개수는 원래 데이터 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 압축효과를 얻을 수 있도록, eigenvector의 개수를 설정할 수 있는 방법이 있는지 설명해보자.**

Eigenvalue의 값이 가장 큰 것이 Variance이 높으므로 eigenvalue가 eigenvector를 선택하는 기준이 되게 된다. 따라서 Eigenvector의 개수를 설정하는 방법으로는 먼저 eigenvalue들을 가장 큰 eigenvalue에 대하여 각각 %화 하여 80% 이상, 90%이상 등의 threshold를 정해 threshold를 넘기면 해당 eigenvalue에 대응하는 eigenvector를 선택하는 방식을 이용하여 eigenvector의 개수를 정할 수 있다.



즉 위의 그래프에서의 예를 보게 되면 80%의 threshold를 넘은 eigenvalue가 3개이므로 eigenvector의 개수도 3개로 설정하고, 80%의 threshold를 넘은 eigenvalue에 대응되는 eigenvector가 선택된다.