4월29일~5월5일

∙ Statistics의 큰 주제를 잡고 공부하였다. 이번주는 Matrix Algebra라는 주제를 가지고 연구하였다.

**2. Matrix Algebra**

이 절에서는 PCA에 필요한 행렬 대수에 대한 배경지식을 제공한다. 특히 주어진 행렬의 고유벡터와 고유 값을 살펴본다.

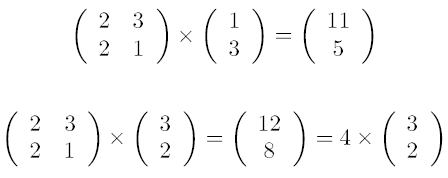
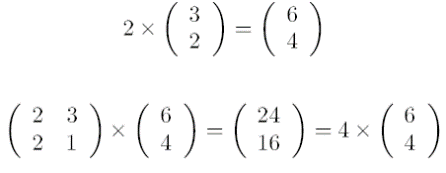
 

Figure 2: Example of how a scaled eigenvector is still and eigenvector

Figure 1: Example of ne non-eigenvector and one eigenvector

**2.1 Eigenvector**

두 개의 행렬을 호환 가능한 크기로 제공하면 두 행렬을 곱할 수 있다. 고유벡터는 특수한 경우이다. Fiqure 1에서 행렬과 벡터 사이의 두 곱셈을 고려할 수 있다.

첫 번째 예제에서 결과 벡터는 원래 벡터의 정수배가 아니지만 두번째 예제에서는 벡터의 4배이다. 왜냐하면 벡터는 2차원 공간의 벡터이기 때문이다. 벡터 (두번째 예제 곱하기에서)는 원점 (0,0)에서 점 (3,2)을 가리키는 화살표를 나타낸다. 다른 행렬인 사각형은 변환 행렬로 생각할 수 있다. 벡터의 왼쪽에 이 행렬을 곱하면 답은 원래 위치에서 변형된 다른 벡터이다.

고유벡터가 발생하는 것은 변환의 본질이다. 왼쪽에 곱해질 때 라인 y=x에서 반사된 벡터를 나타내는 변환 행렬을 상상해봤을 때 라인 y = x 에 놓여있는 벡터가 있다면 그것은 그 자체를 반영한다는 것을 알 수 있다. 이 벡터(벡터의 길이와 상관 없기 때문에 모든 배수)는 해당 변환 행렬의 고유벡터가 된다.

이러한 고유벡터는 어떤 속성을 가질까? 고유벡터가 사각형 행렬에 대해서만 발견될 수 있다는 것을 먼저 알아야한다. 그리고 모든 정사각형 행렬에 고유 벡터가 있는 것은 아니다. 그리고 고유 벡터가 있는 nn행렬이 주어지면 n개의 행렬이 있다. 33행렬에는 3개의 고유 벡터가 있다.

고유벡터의 또 다른 속성은 벡터를 곱하기 전에 어느정도 벡터크기를 조정하더라도 Figure2에서와 같이 결과의 배수가 여전히 동일하다는 것이다. 왜냐하면 어떤 양의 벡터를 스케일링 하면 방향을 바꾸지 않고 더 길게 만들고 있기 때문이다. 마지막으로 행렬의 모든 고유벡터는 수직이다. 많은 차수와는 관계없이 서로 직각을 이룬다. 그래서 x축과 y축으로 표현하는 대신 이러한 수직 고유벡터의 관점에서 데이터를 표현할 수 있다는 것을 의미하므로 중요하다. 나중에 PCA 부분에서 이 작업을 수행할 것이다.

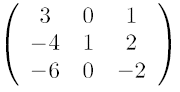
또 다른 중요한 것은 수학자가 고유벡터를 찾을 때 길이가 정확히 1인 고유벡터를 찾는 것 입니다. 알다시피, 벡터의 길이는 그것이 고유벡터 인지 아닌지의 영향을 미치지 않는 반면 방향벡터는 영향을 미치지 않기 때문이다. 따라서 고유 벡터를 표준으로 유지하려면 고유 벡터를 찾을 때마다 일반적으로 길이를 1로 설정하여 모든 고유벡터의 길이를 동일하게 만든다.

는 고유 벡터이고, 그 벡터의 길이는 이다. 그리고 원래의 벡터를 나누어 길이를 1로 만든다.

**2.2 Eigenvalues**

고유 값은 고유벡터와 밀접한 관련이 있다. 실제로 우리는 Fiqure 1에서 고유 값을 보았다. 두 예제 모두에서 정사각형 행렬의 곱한 후 원래 벡터가 스케일 된 양이 같은지 확인했다. 예에서 값은 4이다. 4는 고유벡터와 연관된 고유 값이다. 우리가 정사각형 행렬을 곱하기 전에 우리가 잡은 고유 벡터의 배수가 무엇이든, 항상 4배의 스케일 스케일 벡터를 얻을 것이다 (Figure 2). 따라서 고유 벡터와 고유 값이 항상 쌍을 이루는 것을 볼 수 있다.

다음은 정사각형 행렬에 대해서 예제를 둔다.



다음 벡터 중 어느 것이 해당 행렬의 고유 벡터인지 결정하고 해당 고유 값을 찾아본다.

