Machine Learning

Vector

김선녕(ksycafe@gmail.com)

벡터

고윳값과 고유벡터

벡터의 연산

$\vec{u} = (u_1, u_2, u_3), \qquad \vec{v} = (v_1, v_2, v_3)$

$$u + v = (u_1 + v_1, u_2 + v_2, u_3 + v_3)$$

$$u + v = (u_1 - v_1, u_2 - v_2, u_3 - v_3)$$

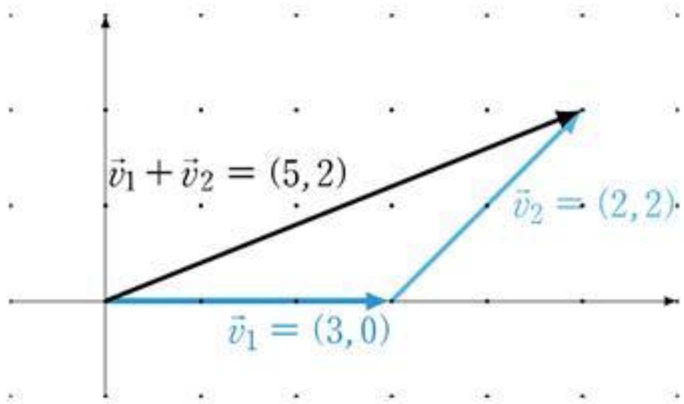
$$\alpha u = (\alpha u_1, \alpha u_2, \alpha u_3)$$

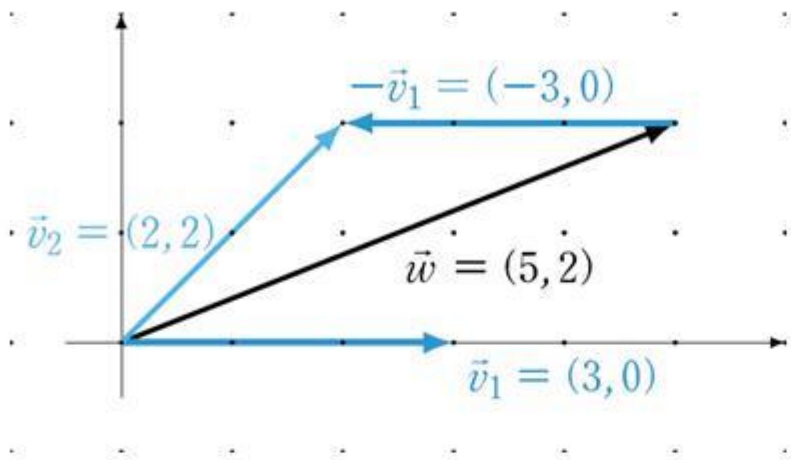
$$u \cdot v = (u_1 v_1 + u_2 v_2 + u_3 v_3)$$

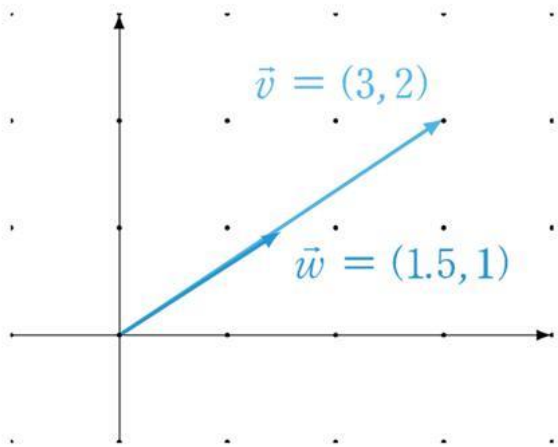
$$u \times v = (u_2 v_3 - u_3 v_2, u_3 v_1 - u_1 v_3, u_1 v_2 - u_2 v_1)$$

$$\|u\| = \sqrt{{u_1}^2 + {u_2}^2 + {u_3}^2}$$

벡터의 덧셈과 뺄셈


$\vec{v}_1 + \vec{v}_2 = (5, 2)$
$\vec{v}_2 = (2, 2)$
$\vec{v}_1 = (3, 0)$


$-\vec{v}_1 = (-3, 0)$
$\vec{v}_2 = (2, 2)$
$\vec{w} = (5, 2)$
$\vec{v}_1 = (3, 0)$


$\vec{v} = (3, 2)$
$\vec{w} = (1.5, 1)$

노름(norm) : 거리측정, 벡터의 크기를 계산하는 함수

- 유클리디안 노름(*Euclidean norm*) : $\boldsymbol{l_2}$노름. $\|\cdot\|_2$ 혹은 $\|\cdot\|$로 표시

$$\|\vec{\boldsymbol{v}}\|_2 = \sqrt{\boldsymbol{v}_1^2 + \boldsymbol{v}_2^2 + \cdots + \boldsymbol{v}_n^2} = \sqrt{\sum_{i=1}^{n} \boldsymbol{v}_i^2}$$

- 맨해튼 노름(*Manhattan Norm*) : $\boldsymbol{l_1}$노름. $\|\cdot\|_1$ 로 표시

$$\|\vec{\boldsymbol{v}}\|_1 = |\boldsymbol{v}_1| + |\boldsymbol{v}_2| + \cdots + |\boldsymbol{v}_n| = \sum_{i=1}^{n} |\boldsymbol{v}_i|$$

- 원소 n개인 벡터 v의 l_p노름

$$\|\vec{\boldsymbol{v}}\|_p = \sqrt[p]{|\boldsymbol{v}_0|^p + |\boldsymbol{v}_1|^p + \cdots + |\boldsymbol{v}_n|^p} = (|\boldsymbol{v}_0|^p + |\boldsymbol{v}_1|^p + \cdots + |\boldsymbol{v}_n|^p)^{\frac{1}{p}}$$

$$= \left(\sum_{i=1}^{n} |\boldsymbol{v}_i|^p\right)^{\frac{1}{p}}$$

  - $\boldsymbol{l_0}$: 단순히 벡터에 있는 0이 아닌 원소의 수
  - $\boldsymbol{p} = \mathbf{1}$ 인 경우 $\boldsymbol{l_1}$노름, $\boldsymbol{p} = \mathbf{2}$ 인 경우 $\boldsymbol{l_2}$노름

- 최대값 노름(*Maximum Norm*) : $\boldsymbol{l_\infty}$ 노름. 벡터에서 가장 큰 절댓값

$$\|\vec{\boldsymbol{v}}\|_\infty = \mathbf{max}(|\boldsymbol{v}_1|, |\boldsymbol{v}_2|, \cdots, |\boldsymbol{v}_n|)$$

노름 예제

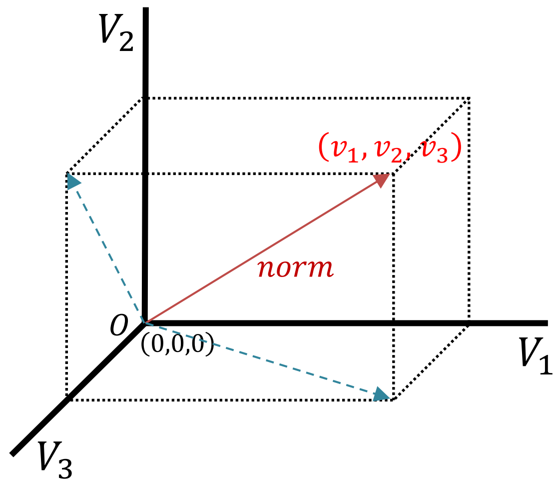
$$v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$$

$$\|\vec{v}\|_2 = (|1|^2 + |-2|^2 + |3|^2)^{\frac{1}{2}} = \sqrt{14}$$ $(p = 2$일때$)$

$$\|\vec{v}\|_2 = \sqrt{v_1^2 + v_2^2 + v_3^2} = \sqrt{1^2 + (-2)^2 + 3^2} = \sqrt{14}$$

$$\|\vec{v}\|_1 = |v_1| + |v_2| + |v_3| = |1| + |-2| + |3| = 6$$

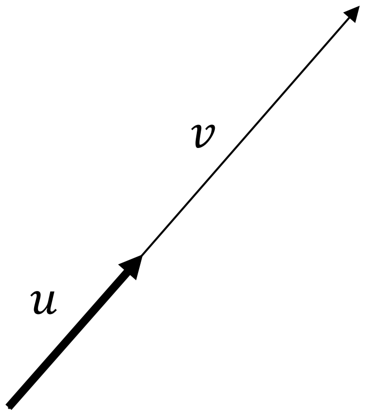
$$\|\vec{v}\|_\infty = \max(|1|, |-2|, |3|) = 3$$



단위벡터(unit vector)

단위벡터 : 노름이 1인 벡터

벡터 v의 단위 벡터 $\boldsymbol{u} = \dfrac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$



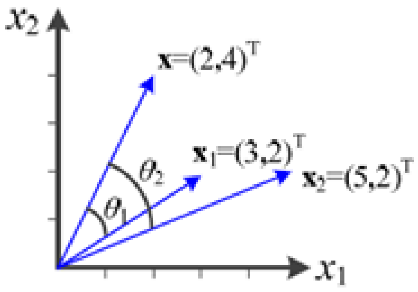
(ex)

$$v = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

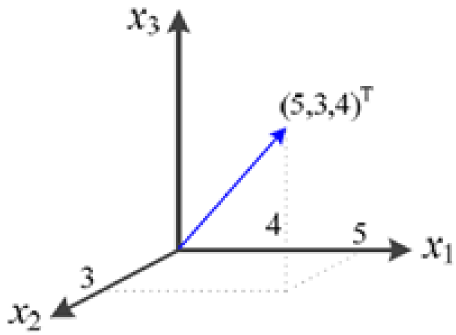
$$\|v\| = \sqrt{1^2 + (2)^2 + 3^2} = \sqrt{14}$$

$$u = \frac{v}{\|v\|} = \begin{bmatrix} \frac{1}{\sqrt{14}} \\ \frac{2}{\sqrt{14}} \\ \frac{3}{\sqrt{14}} \end{bmatrix}$$

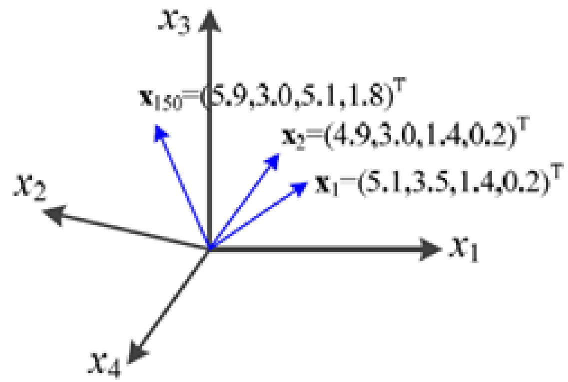
벡터의 기하학적 해석



2차원 벡터



3차원 벡터



4차원 벡터(Iris 데이터)

$$cosine\ similarity(a,b)=\frac{a}{\|a\|}\cdot\frac{b}{\|b\|}=cos(\theta)$$

벡터의 내적(inner product, dot product)

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \qquad v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

$$u \cdot v = u^T \cdot v = [u_1 \quad u_2 \quad \cdots \quad u_n] \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

$$\boldsymbol{v} \cdot \boldsymbol{v} = \boldsymbol{v}^{\boldsymbol{T}} \cdot \boldsymbol{v} = \|\boldsymbol{v}\|^{\boldsymbol{2}}$$

$$v \cdot v = v^T \cdot v = [v_1 \quad v_2 \quad \cdots \quad v_n] \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = v_1 v_1 + v_2 v_2 + \cdots + v_n v_n = \|v\|^2$$

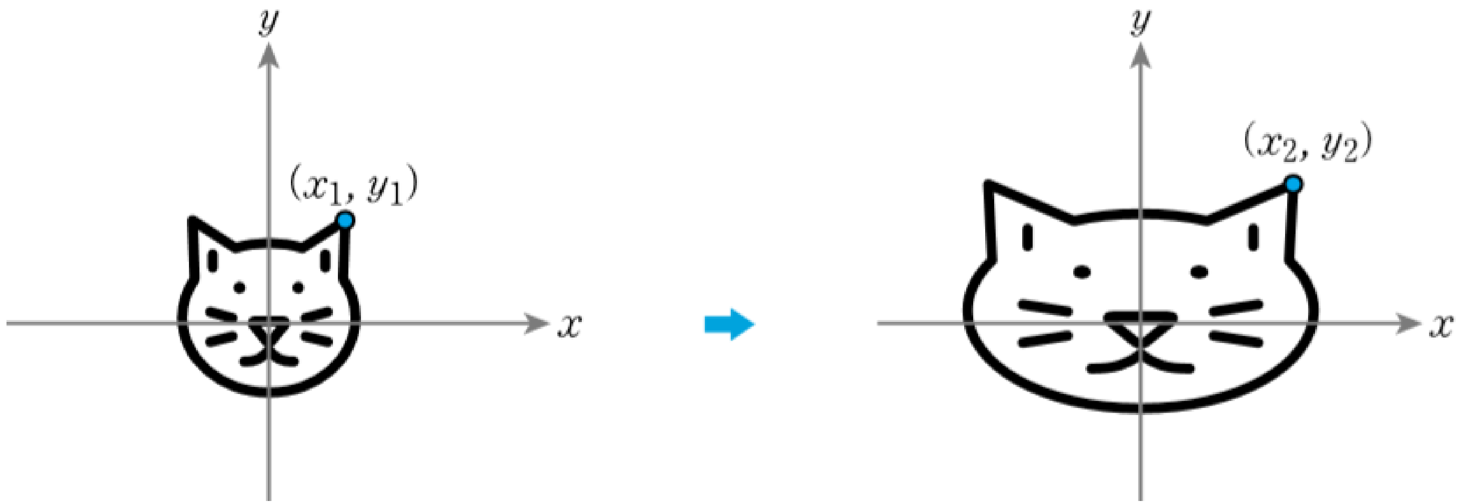
- 벡터의 내적은 교환법칙 성립

  $v_1 \cdot v_2 = v_2 \cdot v_1$

  ${v_1}^T \cdot v_2 = {v_2}^T \cdot v_1$

확대변환과 축소변환

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \qquad \begin{aligned} x_2 &= \alpha x_1 \\ y_2 &= \beta y_1 \end{aligned}$$



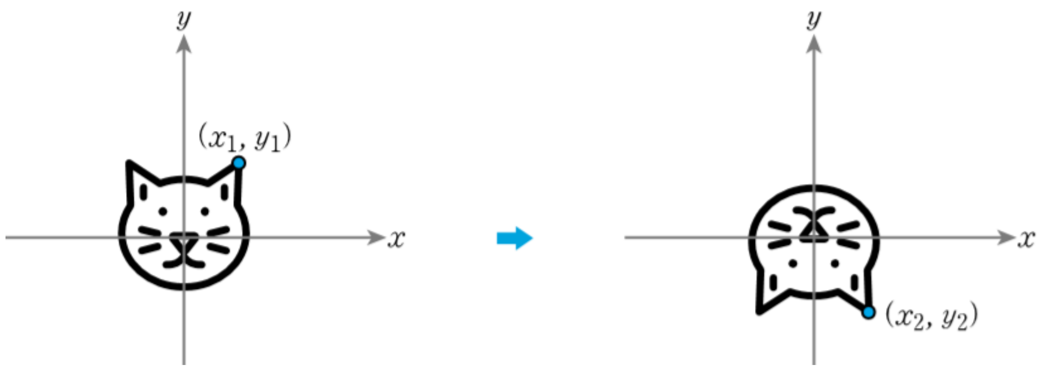
회전변환

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \qquad \begin{aligned} x_2 &= x_1\cos\theta - y_1\sin\theta \\ y_2 &= x_1\sin\theta + y_1\cos\theta \end{aligned}$$

반사변환

- X축 기준으로 반사

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \quad \begin{aligned} x_2 &= x_1 \\ y_2 &= -y_1 \end{aligned} \qquad A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$



- Y축 기준으로 반사

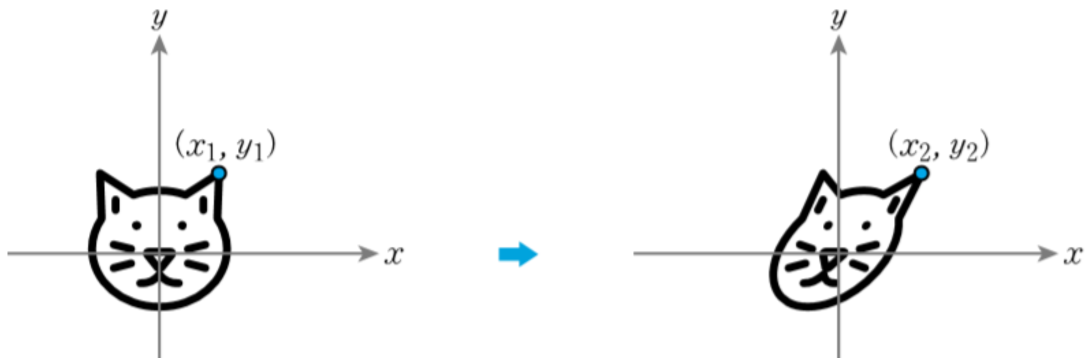
$$B = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

층밀림변환

- x축 방향으로 y의 k배만큼 층밀림

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \quad \begin{aligned} x_2 &= x_1 + ky_1 \\ y_2 &= y_1 \end{aligned}$$

$$A = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$$



- y축 방향으로 x의 k배만큼 층밀림

$$A = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$$

이동변환

$$\begin{cases} x_2 = x_1 + a \\ y_2 = y_1 + b \end{cases}$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} a \\ b \end{bmatrix}$$

벡터

고윳값과 고유벡터

선형독립과 선형종속

벡터공간 V의 벡터 $v_1, v_2, \cdots, v_n$ 에 대하여

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = 0$$

를 만족하는 스칼라 $c_1, c_2, \cdots, c_n$이 $c_1 = c_2 = \cdots = c_n = 0$뿐일 때, 벡터 $v_1, v_2, \cdots, v_n$ 은 **선형독립**이라 한다

벡터공간 V의 벡터 $v_1, v_2, \cdots, v_n$이 선형독립이 아니면 **선형종속**이라 한다

- 선형독립(일차 독립)
  - 한 벡터를 스칼라배해서 다른 벡터를 만들 수 없다면 두 벡터는 일차독립
  - n개의 일차독립인 벡터를 선형결합하면 n차원 공간의 임의의 점을 나타낼 수 있다
- 선형종속(일차 종속)
  - 한 벡터를 스칼라배해서 다른 벡터를 만들 수 있을 때, 두 벡터는 일차종속
  - 일차 종속관계에 있는 벡터는 선형결합으로 오로지 직선밖에 만들지 못한다

- (ex) $\mathbb{R}^3, v_1 = \{1,2,3\}, v_2 = \{-1,0,2\}, v_3 = \{0,2,4\}$

  (sol) $c_1 v_1 + c_2 v_2 + c_3 v_3 = 0 \leftrightarrow (c_1 - c_2, 2c_1 + 2c_3, 3c_1 + 2c_2 + 4c_3) = (0,0,0)$

  $c_1 = c_2 = c_3 = 0$으로 선형독립이다

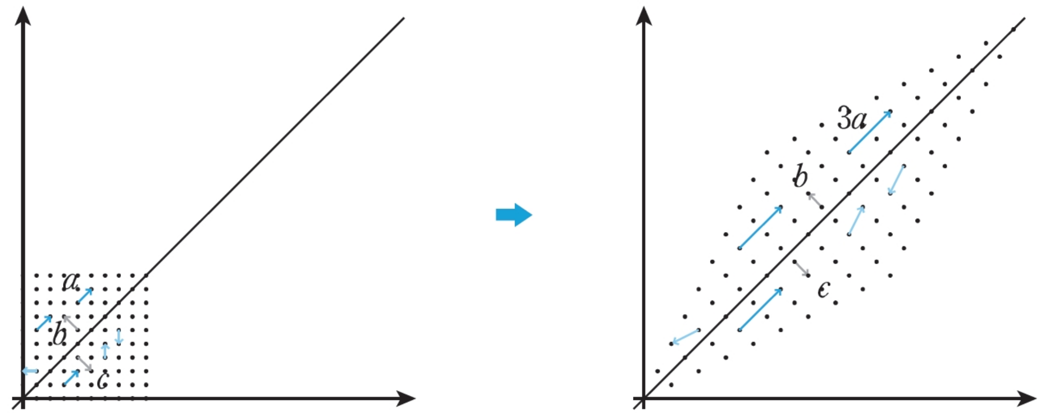
$$\therefore [c_1 \quad c_2 \quad 2c_3] \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 2 \\ 0 & 2 & 4 \end{bmatrix}$$

고윳값과 고유벡터

정방행렬을 사용하여 선형변환할 때 <u>크기는 변하더라도 방향이 변하지 않는 벡터</u>를 **고유벡터**라 하고, 이때 <u>크기 변화의 비율</u>을 **고윳값**이라 한다.

$$Ax = \lambda x \quad \leftrightarrow \quad (A - \lambda I_n)x = 0, \qquad (x \neq 0)$$

- A : n차 정방행렬
- x : 고유벡터$(eigenvector)$
- λ : 고윳값$(eigenvalue)$
- λx : x의 상$(image)$



$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

- 행렬 A와 전치행렬 A^T의 고윳값은 동일하다

고유벡터의 선형독립

- 고유벡터의 선형독립
  - 행렬 $A \in \mathbb{R}^{n \times n}$의 서로 다른 고윳값 $\lambda_1, \lambda_2, \cdots, \lambda_k$에 대응하는 고유벡터 $v_1, v_2, \cdots, v_k$는 서로 선형독립이다

- 대칭행렬
  - 정사각행렬 A가 $A = A^T$를 만족하는 행렬 A

- 벡터의 수직
  - 두 벡터 $u, v \in \mathbb{R}^n$이 $u \cdot v = 0$을 만족하는 경우, 두 벡터 u, v는 서로 수직이다

- 대칭행렬의 고유벡터
  - $A \in \mathbb{R}^{n \times n}$이 대칭행렬일 때, 행렬 A의 서로 다른 고윳값에 대응하는 고유벡터는 서로 수직이다

고윳값, 고유벡터 구하는 예제

$$A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$$

$$\lambda I - A = \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} = \begin{bmatrix} \lambda - 2 & -3 \\ -3 & \lambda + 6 \end{bmatrix}$$

$$def(\lambda I - A) = \begin{bmatrix} \lambda - 2 & -3 \\ -3 & \lambda + 6 \end{bmatrix} = 0 \quad \Rightarrow \quad (\lambda - 2)(\lambda + 6) - 9 = 0$$

$$\Rightarrow \quad \lambda^2 + 4\lambda - 21 = (\lambda - 3)(\lambda + 7) = 0$$

$$\Rightarrow \quad \lambda = \mathbf{3, -7}$$ (고윳값)

① $\lambda_1 = 3$일 때,

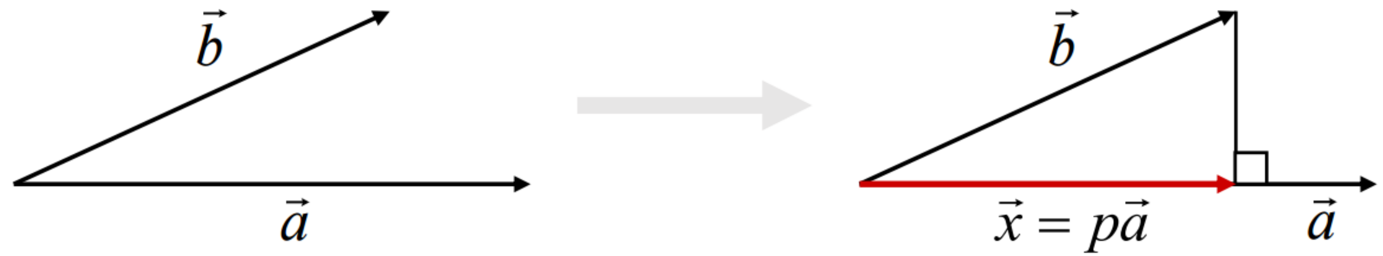
$3I - A = \begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix}$ 이므로, $(3I - A)x = 0$ 에서 고유벡터는 $\begin{bmatrix} \mathbf{3} \\ \mathbf{1} \end{bmatrix}$

② $\lambda_2 = -7$일 때,

$-7I - A = \begin{bmatrix} -9 & -3 \\ -3 & -1 \end{bmatrix}$ 이므로, $(7I - A)x = 0$ 에서 고유벡터는 $\begin{bmatrix} -\frac{\mathbf{1}}{\mathbf{3}} \\ \mathbf{1} \end{bmatrix}$

사영(projection)

- 벡터 b를 벡터 a에 사영한 결과 x



- 벡터 b를 빗변으로 하는 직각 삼각형의 밑변이 벡터 x이면 높이는$\vec{b} - \vec{a}$
- 직교벡터의 내적은 0

$$(\vec{b} - \vec{a})^T \vec{a} = 0$$
$$(\vec{b} - p\vec{a})^T \vec{a} = 0$$
$$\vec{b}^T \vec{a} - p\vec{a}^T \vec{a} = 0$$
$$p = \frac{\vec{b}^T \vec{a}}{\vec{a}^T \vec{a}}$$

a가 유닛벡터($\vec{a}^T \vec{a} = 1$)이면 p는 $\vec{b}^T \vec{a}$, 벡터 a와 b의 내적만으로도 구할 수 있다

공분산 : 두 측정값 사이에 연관성을 분석

n차원의 데이터 $\mathrm{X} = \{x_1, x_2, \cdots, x_n\}$에 대한 평균벡터와 공분산 행렬

m : 평균벡터(*mean vector*), n차원 벡터

C : 공분산 행렬(*covariance matrix*), $n \times n$행렬

$$\boldsymbol{m} = \frac{\boldsymbol{1}}{\boldsymbol{k}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \boldsymbol{x_i} \qquad \boldsymbol{C} = \frac{\boldsymbol{1}}{\boldsymbol{k}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} (\boldsymbol{x_i} - \boldsymbol{m})^{\boldsymbol{T}} (\boldsymbol{x_i} - \boldsymbol{m})$$

$$X^T X = \begin{bmatrix} - & x_1 & - \\ - & x_2 & - \\ & \cdots & \\ - & x_k & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_k \\ | & | & & | \end{bmatrix} = \begin{bmatrix} cov(x_1, x_1) & cov(x_1, x_2) & \cdots & cov(x_1, x_k) \\ cov(x_2, x_1) & cov(x_2, x_2) & \cdots & cov(x_2, x_k) \\ \vdots & \vdots & \ddots & \vdots \\ cov(x_k, x_1) & cov(x_k, x_2) & \cdots & cov(x_k, x_k) \end{bmatrix}$$

$$\frac{X^T X}{k} = \frac{1}{k} \begin{bmatrix} cov(x_1, x_1) & cov(x_1, x_2) & \cdots & cov(x_1, x_k) \\ cov(x_2, x_1) & cov(x_2, x_2) & \cdots & cov(x_2, x_k) \\ \vdots & \vdots & \ddots & \vdots \\ cov(x_k, x_1) & cov(x_k, x_2) & \cdots & cov(x_k, x_k) \end{bmatrix}$$

평균벡터와 공분산 행렬 예제

$$x_1 = \begin{bmatrix}2\\1\end{bmatrix}, \qquad x_w = \begin{bmatrix}2\\3\end{bmatrix}, \qquad x_3 = \begin{bmatrix}4\\1\end{bmatrix}, \qquad x_4 = \begin{bmatrix}4\\5\end{bmatrix}$$

$$m = \frac{1}{k}\sum_{i=1}^{k} x_i = \frac{1}{4}\left(\begin{bmatrix}2\\1\end{bmatrix} + \begin{bmatrix}2\\3\end{bmatrix} + \begin{bmatrix}4\\1\end{bmatrix} + \begin{bmatrix}4\\5\end{bmatrix}\right) = \begin{bmatrix}3\\2.5\end{bmatrix}$$

$$C = \frac{1}{k}\sum_{i=1}^{k}(x_i - m)\,(x_i - m)^T$$

$$= \frac{1}{4}\left(\begin{bmatrix}-1\\-1.5\end{bmatrix}\begin{bmatrix}-1 & -1.5\end{bmatrix} + \begin{bmatrix}-1\\0.5\end{bmatrix}\begin{bmatrix}-1 & 0.5\end{bmatrix} + \begin{bmatrix}1\\-1.5\end{bmatrix}\begin{bmatrix}1 & -1.5\end{bmatrix} + \begin{bmatrix}1\\2.5\end{bmatrix}\begin{bmatrix}1 & 2.5\end{bmatrix}\right)$$

$$= \frac{1}{4}\left(\begin{bmatrix}1 & 1.5\\1.5 & 2.25\end{bmatrix} + \begin{bmatrix}1 & -0.5\\-0.5 & 0.25\end{bmatrix} + \begin{bmatrix}1 & -1.5\\-1.5 & 2.25\end{bmatrix} + \begin{bmatrix}1 & 2.5\\2.5 & 6.25\end{bmatrix}\right)$$

$$= \frac{1}{4}\begin{bmatrix}4 & 2\\2 & 11\end{bmatrix} = \begin{bmatrix}1 & 0.5\\0.5 & 2.75\end{bmatrix}$$

데이터 전처리 - 평균벡터가 영벡터가 되도록 데이터 변환

$$\boldsymbol{x}_i \leftarrow \boldsymbol{x}_i - \boldsymbol{m}$$

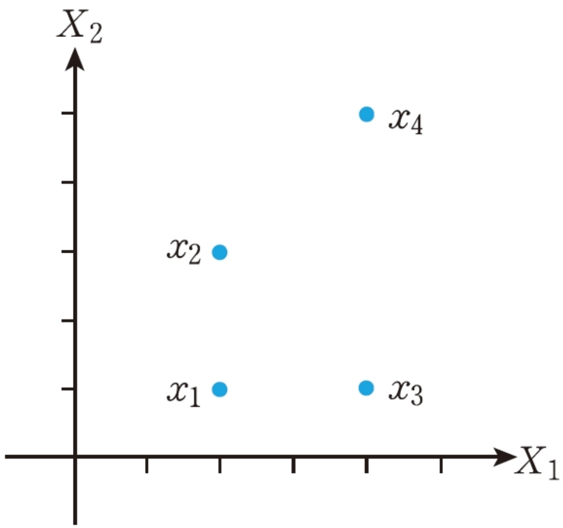
변환된 데이터

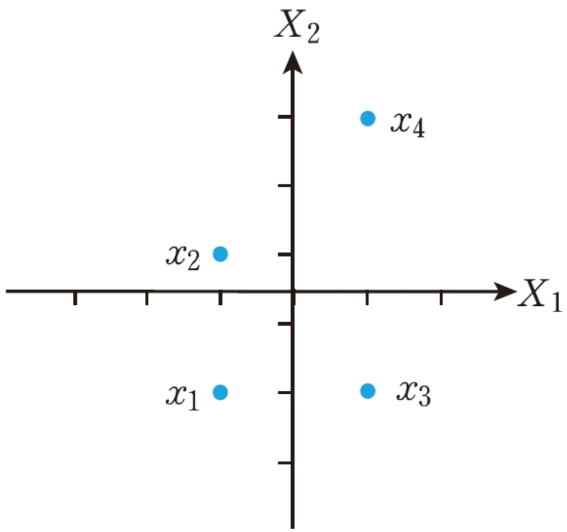
$$x_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2.5 \end{bmatrix} = \begin{bmatrix} -1 \\ -1.5 \end{bmatrix} \qquad x_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ 2.5 \end{bmatrix} = \begin{bmatrix} -1 \\ -0.5 \end{bmatrix}$$

$$x_3 = \begin{bmatrix} 4 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2.5 \end{bmatrix} = \begin{bmatrix} 1 \\ -1.5 \end{bmatrix} \qquad x_2 = \begin{bmatrix} 4 \\ 5 \end{bmatrix} - \begin{bmatrix} 3 \\ 2.5 \end{bmatrix} = \begin{bmatrix} 1 \\ 2.5 \end{bmatrix}$$

변환된 데이터의 평균벡터:영벡터

$$\frac{1}{4}\left(\begin{bmatrix} -1 \\ -1.5 \end{bmatrix} + \begin{bmatrix} -1 \\ -0.5 \end{bmatrix} + \begin{bmatrix} 1 \\ -1.5 \end{bmatrix} + \begin{bmatrix} 1 \\ 2.5 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

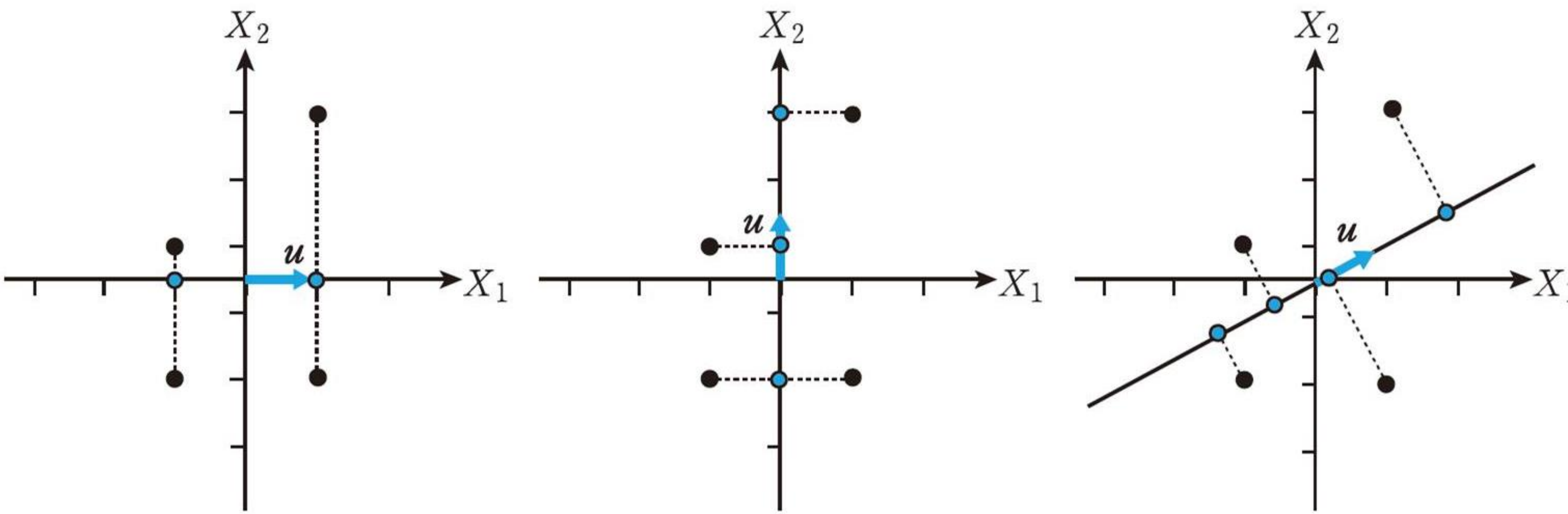




차원축소

- 차원을 줄이면 정보손실(information loss) 발생
- 따라서 차원 축소시에는 정보 손실 최소화

2차원 데이터의 1차원 데이터로의 차원축소의 예



$u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$에 사영

$u = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$에 사영

$u = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$에 사영

분산비교

$$x_1 = \begin{bmatrix}2\\1\end{bmatrix} - \begin{bmatrix}3\\2.5\end{bmatrix} = \begin{bmatrix}-1\\-1.5\end{bmatrix} \qquad x_2 = \begin{bmatrix}2\\3\end{bmatrix} - \begin{bmatrix}3\\2.5\end{bmatrix} = \begin{bmatrix}-1\\-0.5\end{bmatrix}$$

$$x_3 = \begin{bmatrix}4\\1\end{bmatrix} - \begin{bmatrix}3\\2.5\end{bmatrix} = \begin{bmatrix}1\\-1.5\end{bmatrix} \qquad x_2 = \begin{bmatrix}4\\5\end{bmatrix} - \begin{bmatrix}3\\2.5\end{bmatrix} = \begin{bmatrix}1\\2.5\end{bmatrix}$$

1. $u = [1 \quad 0]^T$, 평균분산 $= \mathbf{1.0}$

$$[1 \quad 0]\begin{bmatrix}-1\\-1.5\end{bmatrix} = -1, [1 \quad 0]\begin{bmatrix}-1\\-0.5\end{bmatrix} = -1, [1 \quad 0]\begin{bmatrix}1\\-1.5\end{bmatrix} = 1, [1 \quad 0]\begin{bmatrix}1\\2.5\end{bmatrix} = 1$$

2. $u = \begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}^T$, 평균분산 $= \mathbf{1.22}$

$$\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}-1\\-1.5\end{bmatrix} = -\frac{2.5}{\sqrt{2}}, \qquad \begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}-1\\-0.5\end{bmatrix} = -\frac{1.5}{\sqrt{2}}$$

$$\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}1\\-1.5\end{bmatrix} = -\frac{0.5}{\sqrt{2}}, \qquad \begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}1\\2.5\end{bmatrix} = -\frac{3.5}{\sqrt{2}}$$

- n차원 데이터 $\{x_1, x_2, \cdots, x_m\}$을 새로운 기저 $\{u_1, u_2, \cdots, u_n\}$의 좌표계로 선형변환할 때, 임의의 n차원 데이터 x (u에서 $x^T u$만큼의 거리)

$$x = \sum_{i=1}^{n} (x^T u_i) u_i \qquad (x^T u_i = x \cdot u_i \text{는 } x\text{를 } u_i\text{방향으로 정사영한 벡터의 크기})$$

n차원 좌표계에서 K개의 기저벡터로 x를 으로 근사 : $\hat{x} = \sum_{i=1}^{K} (x^T u_i) u_i$

$$J(\text{오차}) = \frac{1}{n}\sum_{i=1}^{n} (x_i - \hat{x}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} ({x_i}^T u_j) u_j - \sum_{j=1}^{K} ({x_i}^T u_j) u_j\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=K+1}^{n} ({x_i}^T u_j) u_j\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=K+1}^{n} \left({x_i}^T u_j\right)^2 \qquad (v^T v = v^2)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=K+1}^{n} \left({x_i}^T u_j\right)^T \left({x_i}^T u_j\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=K+1}^{n} {u_j}^T x_i {x_i}^T u_j$$

$$= \sum_{j=K+1}^{n} {u_j}^T \left(\frac{1}{n}\sum_{i=1}^{n} x_i {x_i}^T\right) u_j$$

$$= \sum_{j=K+1}^{n} {u_j}^T C u_j \qquad \left(\text{평균벡터가 영벡터이므로 } \frac{1}{n}\sum_{i=1}^{n} x_i {x_i}^T = C\right)$$

목적함수

$$u_j{}^T C u_j = u_j{}^T \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i{}^T \right) u_j = \frac{1}{n} \sum_{i=1}^{n} u_j{}^T x_i x_i{}^T u_j = \frac{1}{n} \sum_{i=1}^{n} \left(x_i{}^T u_j \right)^T \left(x_i{}^T u_j \right) = \sigma_j{}^2$$

분산($u_j{}^T C u_j = \sigma_j{}^2$)을 가장 크게 하는 기저벡터 u를 찾는 최적화 문제
(제한조건 : $u^T u = 1$)

라그랑주 승수 λ를 사용하여 목적함수 $\mathcal{L}$정의

$$\mathcal{L} = u^T C u - \lambda (u^T u - 1)$$

$$\frac{\delta \mathcal{L}}{\delta u^T} = Cu - \lambda u = 0 \quad \Longrightarrow \quad \boldsymbol{Cu = \lambda u}$$

(u는 공분산 행렬 C의 고유벡터, λ는 고윳값)

$$u^T C u = \lambda u^T u = \lambda = \sigma^2$$

데이터의 공분산 행렬 C로 부터 고윳값(λ)이 큰 순서대로 K의 고유벡터
$\{u_1, u_2, \cdots, u_K\}$ 를 선택

$$\{u_1, u_2, \cdots, u_K\} \;\; (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K)$$

이때 선택된 축을 주성분 축(*principal axis*)혹은 주성분 벡터(*principal vector*)
라 한다.

주성분 분석 예제

$$x_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad x_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad x_3 = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \quad x_4 = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

$$m = \begin{bmatrix} 3 \\ 2.5 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2.75 \end{bmatrix}$$

$$\lambda_1 = \frac{15 + \sqrt{65}}{8}, \qquad u_1 = \begin{bmatrix} \frac{\sqrt{65} - 7}{4} \\ 1 \end{bmatrix} \approx \begin{bmatrix} 0.27 \\ 1 \end{bmatrix}$$

$$\lambda_2 = \frac{15 - \sqrt{65}}{8}, \qquad u_2 = \begin{bmatrix} \frac{\sqrt{65} + 7}{4} \\ 1 \end{bmatrix} \approx \begin{bmatrix} -3.77 \\ 1 \end{bmatrix}$$

$\lambda_1 > \lambda_2$이므로 u_1 사용

$$\hat{x}_1 \approx {x_1}^T u_1 = \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} 0.27 \\ 1 \end{bmatrix} = 1.54$$

$$\hat{x}_2 \approx {x_2}^T u_1 = \begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} 0.27 \\ 1 \end{bmatrix} = 3.54$$

$$\hat{x}_3 \approx {x_3}^T u_1 = \begin{bmatrix} 4 & 1 \end{bmatrix} \begin{bmatrix} 0.27 \\ 1 \end{bmatrix} = 2.08$$

$$\hat{x}_4 \approx {x_4}^T u_1 = \begin{bmatrix} 4 & 5 \end{bmatrix} \begin{bmatrix} 0.27 \\ 1 \end{bmatrix} = 6.08$$

특잇값(singular value)

- $m \times n$ 실수행렬 A의 특잇값은 A^TA에 대한 고윳값의 양의 제곱근이다. 즉, A 의 특잇값 σ_i는 A^TA 에 대한 고윳값 λ_i의 양의 제곱근 $\sqrt{\lambda_i}$ 이다.

[예] $A = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$의 특이값은?

$$A^TA = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}$$

$$\det(\lambda I - A^TA) = \lambda(\lambda - 10)(\lambda - 12) = 0$$

A^TA의 고윳값 : $\lambda_1 = 12, \quad \lambda_2 = 10, \quad \lambda_3 = 0$

A의 특잇값 : $\sigma_1 = 2\sqrt{3}, \quad \sigma_2 = \sqrt{10}, \quad \sigma_3 = 0$

- $m \times n$ 실수행렬 A를 $A = U\Sigma V^T$로 표현한 것을 **특잇값 분해**($\boldsymbol{SVD}$)라고 한다. 이때 U는 $m \times m$ 직교행렬, V는 $n \times n$ 직교행렬, Σ는 주대각 성분에 A의 특잇값을 큰 것부터 차례대로 넣은 $m \times n$ 직사각 대각행렬이다
  - 직사각 대각행렬 ($rectangular\ diagonal\ matrix$): 주대각 성분에 특잇값이 위치하고 나머지 성분은 모두 0인 직사각행렬(Σ)

$$A = U\Sigma V^T$$
$$\Downarrow$$
$$\begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \sqrt{0.2} & 0 & 0 & 0 & \sqrt{0.8} \\ 0 & 0 & 0 & 1 & 0 \\ -\sqrt{0.8} & 0 & 0 & 0 & \sqrt{0.2} \end{bmatrix}$$

U는 4×4직교행렬

Σ는 A의 특잇값을 주대각 성분으로 갖는 4×5직사각 대각행렬

V는 5×5직교행렬.

특이값 분해의 행렬 전개

$$A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i {v_i}^T \qquad (r\text{은 } A\text{의 계수. 특잇값 개수})$$

$$\begin{bmatrix} \vdots & \vdots & & \vdots \\ u_1 & u_2 & \cdots & u_m \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & \cdots & 0 \\ & \ddots & & \vdots & \ddots & \vdots \\ & & \sigma_r & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \cdots & v_1 & \cdots \\ \cdots & v_2 & \cdots \\ & \vdots & \\ \cdots & v_n & \cdots \end{bmatrix}$$

$$= \sigma_1 u_1 {v_1}^T + \cdots + \sigma_r u_r {v_r}^T + 0 + \cdots + 0$$

특잇값 분해를 이용한 주성분 분석

$$X = U\Sigma V^T$$

평균벡터가 영벡터라 가정하면 공분산 행렬 C는

$$\begin{aligned} C &= \frac{1}{m-1}X^TX = \frac{1}{m-1}(U\Sigma V^T)^T U\Sigma V^T \\ &= \frac{1}{m-1}V\Sigma^T U^T U\Sigma V^T \\ &= \frac{1}{m-1}V\Sigma^T\Sigma V^T \\ &= VSV^T \qquad \left(S = \frac{1}{m-1}\Sigma^T\Sigma\right) \end{aligned}$$

따라서 $C = VSV^T$는 C의 고윳값 분해에 해당.

그러므로 C의 고유벡터인 V의 열벡터는 주성분 벡터이고 V의 고윳값 λ_i는 특이값 σ_i에 대해 $\lambda_i = \sigma_i{}^2 \,/\, (m-1)$이다.

$X = U\Sigma V^T$에서 V의 각 열벡터는 주성분 벡터. 주성분 분석에서 V에서 왼쪽에 있는 k개의 열벡터를 주성분 벡터로 사용

차원축소 : xV_k $\left(V_k\text{는 } V\text{의 왼쪽} k\text{개의 열벡터}\right)$

특잇값 분해를 이용한 주성분 분석 예제

$$x_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \qquad x_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \qquad x_3 = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \qquad x_4 = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

X' : 평균벡터(m)가 영벡터가 되도록 각 데이터에서 평균벡터를 뺀 행렬

$$X = \begin{bmatrix} 2 & 1 \\ 2 & 3 \\ 4 & 1 \\ 4 & 5 \end{bmatrix} \qquad m = \begin{bmatrix} 3 & 2.5 \end{bmatrix} \qquad X' = \begin{bmatrix} -1 & -1.5 \\ -1 & 0.5 \\ 1 & -1.5 \\ 1 & 2.5 \end{bmatrix}$$

특이값 분해

$$U \approx \begin{bmatrix} -0.50 & -0.31 & 0.08 & 0.80 \\ 0.07 & -0.59 & 0.76 & -0.26 \\ -0.35 & 0.73 & 0.59 & 0.01 \\ 0.79 & 0.17 & 0.25 & 0.54 \end{bmatrix} \qquad \Sigma \approx \begin{bmatrix} 3.40 & 0 \\ 0 & 1.86 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad V \approx \begin{bmatrix} \mathbf{0.26} & 0.97 \\ \mathbf{0.97} & -0.26 \end{bmatrix}$$

V의 첫 번째 열벡터를 주성분 벡터 v로 선택 : $v = \begin{bmatrix} 0.26 \\ 0.97 \end{bmatrix}$

$$\hat{x}_1 \approx {x_1}^T v = \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{0.26} \\ \mathbf{0.97} \end{bmatrix} = 1.49 \qquad \hat{x}_2 \approx {x_2}^T v = \begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} \mathbf{0.26} \\ \mathbf{0.97} \end{bmatrix} = 3.43$$

$$\hat{x}_3 \approx {x_3}^T v = \begin{bmatrix} 4 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{0.26} \\ \mathbf{0.97} \end{bmatrix} = 2.01 \qquad \hat{x}_4 \approx {x_4}^T v = \begin{bmatrix} 4 & 5 \end{bmatrix} \begin{bmatrix} \mathbf{0.26} \\ \mathbf{0.97} \end{bmatrix} = 5.89$$