행렬연산의 기초

Jinseog Kim

October 8, 2007

1 행렬연산의 기초

1.1 행렬의 종류

- 정방행렬(square matrix) 행수와 열수가 같은 행렬 $A_{n\times n}$
- 대각행렬(diagonal matrix) 대각선원소를 제외한 모든 원소가 0인 행렬
- 항등행렬(Identity matrix): 대각행렬 중 대각선 원소가 모두 1인 행렬

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

- 정방 행렬의 원소가 모두 1인 행렬

$$J = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 1 \end{pmatrix} = \mathbf{1}\mathbf{1}'$$

- 전치행렬(Transpose matrix): 행렬 A의 원소 a_{ij}의 행번호와 열번호를 바꾼 행렬.

$$A' = A^\top = (a_{ji})$$

- 대칭행렬(Symmetric matrix):

$$A' = A.$$

- 역행렬(Inverse matrix): 행렬 A에 대하여, 어떤 행렬 B존재하여, 다음을 만족할 때, B를 역행렬이라고 하고 $B = A^{-1}$로 표현한다.

$$AB = BA = I.$$

- 직교행렬(Orthogonal matrix):

$$AA' = A'A = I, \text{혹은 } A' = A^{-1}.$$

1.2 선형변환

행렬 A, 열벡터 $\mathbf{b}$,

$$\mathbf{y} = A\mathbf{x} + \mathbf{b}$$

를 $\mathbf{x}$의 $\mathbf{y}$로의 선형변환(linear transformation)이라고 한다.

- If $E(\mathbf{x}) = \boldsymbol{\mu}$, $Cov(\mathbf{x}) = \Sigma_{\mathbf{x}}$, then

$$E(\mathbf{y}) = A\boldsymbol{\mu}_x + \mathbf{b},\ Cov(\mathbf{y}) = A\Sigma_{\mathbf{x}}A^{\top}.$$

- **Affine transformation**
  만일 A가 non-singular matrix(정칙행렬, 역행렬이 존재하는 경우)이면,
  $\mathbf{y} = A\mathbf{x} + \mathbf{b}$를 Affine transformation이라고 부른다.

1.3 선형독립(linearly independent)

- n개의 열벡터 $\mathbf{x}_1, \ldots, \mathbf{x}_n$가 0이 아닌 상수($a_i$)와의 선형결합

$$\mathbf{a}'\mathbf{x} = a_1\mathbf{x}_1 + \ldots + a_n\mathbf{x}_n = 0$$

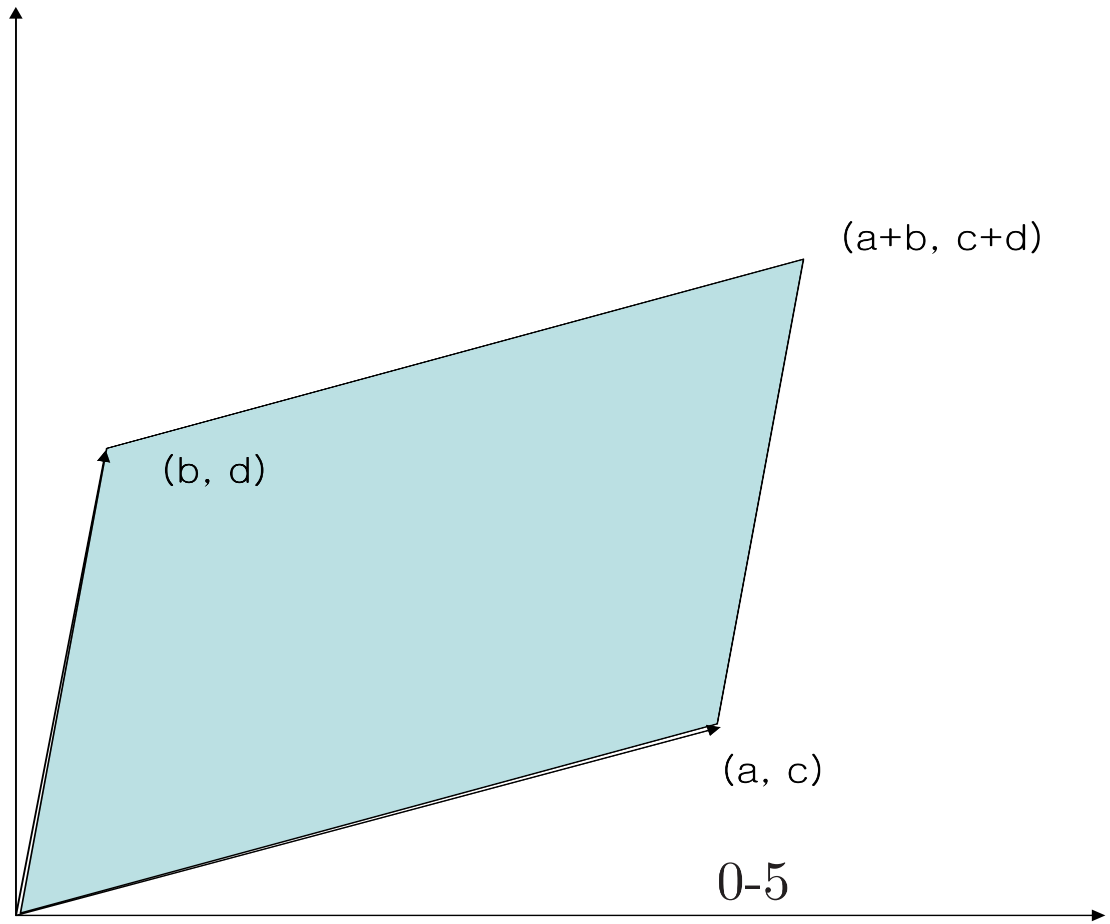
  을 만족하면 $\mathbf{x}_1, \ldots, \mathbf{x}_n$를 선형종속(linearly dependent)라고 부르며, $\mathbf{a} \ = \mathbf{0}$이면 선형독립이라고 부른다.

- 행렬 X는 ,열벡터들 $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$으로 이루어 졌다고 할때, 선형독립인 열의 개수를 행렬 X의 계수(rank)라고 한다.

1.4 행렬식(determinant)

행렬식(determinant)는 정방 행렬 $A_{n\times n}$를 scalar값으로 mapping(대응)시킴. (det(A) 혹은 $|A|$로 표현)

의미: 행렬 A를 n차원 공간에서의 n개의 벡터라고 할 때, 그 벡터들로 만들어지는 도형의 부피(Volumn)의 개념으로 이해할 수 있다.


(a+b, c+d)
(b, d)
(a, c)
0-5

1.5 trace

정방행렬의 대각선 원소의 합을 행렬의 trace라고 한다.

$$tr(A+B) = tr(A) + tr(B)$$
$$tr(aA) = a \times tr(A)$$
$$tr(A) = tr(A')$$

.

If $A_{n\times n}$ matrix and $B_{n\times n}$ matrix, then

$$tr(AB) = tr(BA).$$

$$\mathrm{tr}(AB) = \sum_{i=1}^{n}(AB)_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}B_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{n}B_{ji}A_{ij} = \sum_{j=1}^{n}(BA)_{jj} = \mathrm{tr}(BA)$$

$$tr(ABC) = tr(CAB) = tr(BCA)$$

1.6 이차형식(Quadratic form)

$A_{p\times p}$ 가 대칭행렬(symmetric)이고 $\mathbf{x}$ 가 p 열벡터 일때,

$$Q(\mathbf{x}) = \mathbf{x}'A\mathbf{x}$$

를 $\mathbf{x}$의 A에 대한 이차형식이라고 부른다.

- 만일 $\mathbf{x}'A\mathbf{x} > 0$ (resp. $\mathbf{x}'A\mathbf{x} < 0$) for every vector $\mathbf{x} \neq 0$ 이면 A를 "양정치(positive definite)" (resp. 음정치(negative definite))행렬 이라고 부른다.

- If we change the strict inequality into $\geq ; \leq$, A를 "semi-definite"(즉, 양반정치, 음반정치) 행렬 이라고 부른다.

- If $Q(v) < 0$ for some v and $Q(v) > 0$ for some other v, Q is said to be "indefinite".

이차형식은 χ^2 분포, mahalanobis 거리와 같은 데서 이용된다.

1.7 고유치, 고유벡터-eigen value and eigen vector

주어진 행렬 A 에 대하여 어떤 벡터($\mathbf{x}$)의 선형변환이 자기 자신($\mathbf{x}$)과 어떤 상수(λ)의 곱으로 표현할 수 있다고 하자. 즉,

$$A\mathbf{x} = \lambda\mathbf{x}.$$

이는

$$A\mathbf{x} = (\lambda I)\mathbf{x},$$

또는

$$(A - \lambda I)\mathbf{x} = 0$$

로 표현될 수 있다.

이 때, eigenvector는 위 식을 만족하는 **0**이 아닌 벡터, eigenvalue는 역시 0이 아닌 스칼라 값이다. 다음의 예를 살펴보자.

$$A = \begin{bmatrix} 0 & 0 \\ -\frac{1}{2} & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1-\lambda & 0 \\ -\frac{1}{2} & 1-\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

여기서 eigen values와 eigen-vector는 $\lambda = 1, \mathbf{x} = \begin{bmatrix} 0 \\ c \end{bmatrix}$

See page 43–46 for the properties of eigen-value and eigen vectors.

1.8 멱등행렬(idempotent matrix)

$$A^2 = AA = A.$$

Example: 회귀분석에서 hat matrix $H = X(X'X)^{-1}X'$.

종속변수의 관측치를 행렬로 표현하면 $\mathbf{y}$이고, 회귀계수의 추정치는 $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$이다. 이 때, $\mathbf{y}$의 추정치는

$$\hat{\mathbf{y}} = X\hat{\boldsymbol{\beta}} = X(X'X)^{-1}X'\mathbf{y} = H\mathbf{y}.$$

또한 잔차벡터는

$$\mathbf{y} - X\hat{\boldsymbol{\beta}} = [I - X(X'X)^{-1}X']\mathbf{y} = (I - H)\mathbf{y}.$$

$$H^2 = X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X' = H.$$

이므로 $X(X'X)^{-1}X'$는 멱등행렬이다.

1.9 Spectral decomposition

Let A be a positive definite(pd) matrix. Then there exists a orthogonal matrix P, and diagonal matrix Λ,

$$A = P\Lambda P'.$$

If diagonal elements of Λ are the eigen values of A, then P is composed of the eigen-vectors, so

$$A = P\Lambda P'.$$

1.10 다변량자료의 산포 및 거리

다변량자료에서의 자료점들의 흩어져 있는 정도를 공분산행렬($S(\Sigma)$)로 나타낸다. 이는 행렬로 표현되어 있어서 쉽게 이해하기가 어렵다. 따라서 하나의 스칼라값으로 표현하여 산포의 정도를 쉽게 나타낼 수 있다. 행렬의 행렬식(determinant)는 행렬에 포함되 있는 벡터들로 나타내어 지는 공간의 부피라고 한 적이 있다. 이를 공분산행렬에 적용하여 공분산행렬의 부피를 계산한 것이 일반화된 분산(generalized variance)이다.

$$|S| = \prod_{i=1}^{p} l_i$$

자료간의 거리는 ?

- Euclidian distance

$$\sqrt{(\mathbf{x}-\mathbf{y})'(\mathbf{x}-\mathbf{y})}$$

- Mahalanovis distance

$$(\mathbf{x}-\mathbf{y})'S^{-1}(\mathbf{x}-\mathbf{y})$$