Linear Algebra for Statistics

Chapter 6

Instructor: Seokho Lee (이석호)

Hankuk University of Foreign Studies

Chapter 6 행렬의 응용

6.1 마르코프 연쇄와 전이행렬

확률과정(stochastic process)이란 서로 다른 n 개(혹은 무한개)의 상태(state)가 있을 때, 한 상태에서 다른 상태로 옮겨지는 확률적 과정을 의미한다.

확률과정에서 한 상태에서 다른 상태로 이동(전이, transition)할 확률을 일단계 전이확률 (one-step transition probability)라하며, 이를 $n \times n$ 행렬로 표현한 것을 전이행렬(transition matrix)이라 한다.

j 번째 상태에서 i 번째 상태로 전이되는 확률을

$$p_{ij} = \Pr(X_t = i | X_{t-1} = j)$$

라고 하면, 전이행렬은 아래와 같다.

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}. \tag{6.1}$$

여기서, 열(column)은 '출발'을, 행(row)은 '도착'을 나타내며, 행렬의 (i, j) 원소 p_{ij} 는 j 상태에서 출방하여 i 상태로 전이될 확률을 나타낸다.

Definition

행렬 A 가 전이행렬이 되기 위해서 아래의 조건을 만족하여야 한다.

1. A 는 $n \times n$ 정방행렬이어야 한다.
2. A 의 모든 원소는 $0 \leq a_{ij} \leq 1$ 이어야 한다.
3. A 의 모든 열에 대하여 열방향 원소의 합은 항상 1이다. 즉, $\sum_{i=1}^{n} p_{ij} = 1$, $\forall j = 1, 2, \ldots, n$.

Example

예를 들어 다음의 행렬은 전이행렬의 조건을 만족시킨다.

$$P = \begin{pmatrix} 0.9 & 0.15 \\ 0.1 & 0.85 \end{pmatrix} \tag{6.2}$$

Definition

어떤 상태로 이동하는 확률이 바로 이전 시점의 상태가 어떤 것인지에 영향을 받는 경우, 이를 마르코프 연쇄(Markov Chain)라고 하며, 이 연쇄를 설명하는 확률행렬을 전이행렬 (transition matrix)이라 한다.

Example

다음의 전이행렬

$$P = \begin{pmatrix} 0.9 & 0.15 \\ 0.1 & 0.85 \end{pmatrix}$$

이 마르코스 연쇄를 설명한다고 하자. 현재 시점에서 $\mathbf{x}_t = (20, 70)^T$ 이라고 한다면 다음 시점에서의 값을 $\mathbf{x}_{t+1}$ 라고 하면 마르코프 연쇄에 의해

$$\begin{aligned} \mathbf{x}_{t+1} &= P\mathbf{x}_t && (6.3) \\ &= \begin{pmatrix} 0.9 & 0.15 \\ 0.1 & 0.85 \end{pmatrix} \begin{pmatrix} 20 \\ 70 \end{pmatrix} && (6.4) \\ &= \begin{pmatrix} 28.5 \\ 61.5 \end{pmatrix} && (6.5) \end{aligned}$$

Example (continue)

이러한 추세로 그 다음 시점의 값 $\mathbf{x}_{t+2}$를 예측하면 마르코프 연쇄에 의해

$$\begin{aligned}
\mathbf{x}_{t+2} &= PP\mathbf{x}_t = P\mathbf{x}_{t+1} && (6.6)\\
&= \begin{pmatrix} 0.9 & 0.15 \\ 0.1 & 0.85 \end{pmatrix}^2 \begin{pmatrix} 20 \\ 70 \end{pmatrix} = \begin{pmatrix} 0.9 & 0.15 \\ 0.1 & 0.85 \end{pmatrix} \begin{pmatrix} 28.5 \\ 61.5 \end{pmatrix} && (6.7)\\
&= \begin{pmatrix} 34.9 \\ 55.1 \end{pmatrix} && (6.8)
\end{aligned}$$

가 된다.

Example

택시들이 도시1, 도시2, 도시3 을 돌아다니며 운행하고 있다. 각 도시에서 손님을 태운다고 가정했을 때, 택시에 탄 손님이 각 도시를 목적지로 할 확률은 그림(Figure 6.1)과 같다. 각 도시에서 다른 도시에 이르는 확률을 전이행렬로 표시하라.

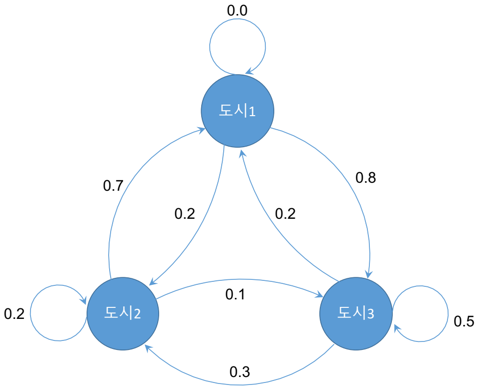

Figure 6.1: 택시 운전사가 도시 사이를 운행하면서 각 도시방향의 손님을 태울 확률

Example (continue)

(풀이) 전이행렬은 다음과 같다.

$$P = \begin{pmatrix} 0.0 & 0.7 & 0.2 \\ 0.2 & 0.2 & 0.3 \\ 0.8 & 0.1 & 0.5 \end{pmatrix} \tag{6.9}$$

Example

택시를 기다리는 사람의 수가 (도시1, 도시2, 도시3) = (10, 20, 30)이라고 할 때, 모든 사람이 택시를 타고 난 후 다음 도시에서 택시를 기다린다고 가정할 때, 두번째 택시를 기다리는 사람의 수는 얼마로 예상하는가?

(풀이) $\mathbf{x} = (10, 20, 30)^T$ 이라 하면,

$$P\mathbf{x} = \begin{pmatrix} 0.0 & 0.7 & 0.2 \\ 0.2 & 0.2 & 0.3 \\ 0.8 & 0.1 & 0.5 \end{pmatrix} \begin{pmatrix} 10 \\ 20 \\ 30 \end{pmatrix} = \begin{pmatrix} 20 \\ 15 \\ 25 \end{pmatrix}$$

6.2 선형함수와 선형변환

Definition

함수 $f(x)$ 가 아래 조건을 만족할 때

1. 가법성(additivity) : $f(x_1 + x_2) = f(x_1) + f(x_2)$
2. 동질성(homogeneity) : $f(ax) = af(x)$, a 는 상수

이러한 $f(x)$ 를 선형함수(linear function)라고 한다.

Remark 위의 조건 하에서 다음을 만족한다

$$f(a_1x_1 + a_2x_2) = a_1f(x_1) + a_2f(x_2). \tag{6.10}$$

Example

함수 $f(x)$ 가 선형함수인지 아닌지 판단하라.

(1) $f(x) = 2x$

(2) $f(x) = x^2$

(3) $f(x) = \ln x$

(4) $f(x) = 1 + x$

(풀이) $f(x) = 2x$ 는

$$f(a_1x_1 + a_2x_2) = 2(a_1x_1 + a_2x_2) = a_1(2x_1) + a_2(2x_2) = a_1f(x_1) + a_2f(x_2)$$

이므로 선형함수이다. 다른 함수들은 선형함수가 아님을 손쉽게 보일 수 있다. (4)의 경우, 직선함수이나 선형함수가 아님에 유의할 것.

Definition

두 벡터 $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$ 와 행렬 $A_{m \times n}$ 에 대하여, 함수 $f : \mathbb{R}^n \to \mathbb{R}^m$ 가

$$\mathbf{y} = f(\mathbf{x}) = A\mathbf{x} \tag{6.11}$$

라고 할 때, f를 선형변환이라 한다.

Proposition (6.1)

선형변환 $\mathbf{y} = A\mathbf{x}$ 는 선형함수임을 보여라.

Proof.

$$\begin{aligned} f(a_1\mathbf{x}_1 + a_2\mathbf{x}_2) &= A(a_1\mathbf{x}_1 + a_2\mathbf{x}_2) && (6.12) \\ &= a_1A\mathbf{x}_1 + a_2A\mathbf{x}_2 && (6.13) \\ &= a_1f(\mathbf{x}_1) + a_2f(\mathbf{x}_2) && (6.14) \end{aligned}$$

이므로 선형함수이다. □

6.3 선형변환의 종류

6.3.1 회전변환

회전변환(rotation)은 공간상의 점들을 반시계방향(counter-clockwise)으로 θ 만큼 회전시킨 변환을 의미한다.

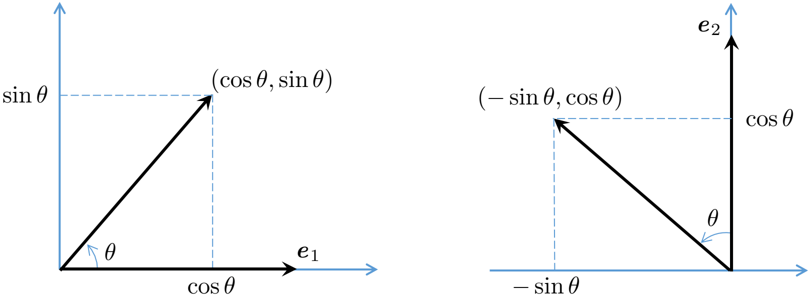

Figure 6.2: 회전변환에 의해 변환된 벡터

이차원공간 상에서 점들을 θ 만큼 반시계방향으로 회전하는 변환에 대응하는 선형변환을

$$\mathbf{y} = A\mathbf{x} \qquad \text{with} \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

라고 하자.

Figure 6.2의 왼쪽 그림에 의하면, 기본벡터 $\mathbf{e}_1 = (1, 0)^T$ 을 반시계방향으로 θ 만큼 회전하여 얻은 벡터는 $(\cos\theta, \sin\theta)$ 이므로 다음이 성립한다.

$$\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix}. \tag{6.15}$$

즉,

$$a = \cos\theta, \qquad c = \sin\theta$$

이다.

Figure 6.2의 오른쪽 그림에 의하면, 기본벡터 $\mathbf{e}_2 = (0,1)^T$ 을 반시계방향으로 θ 만큼 회전하여 얻은 벡터는 $(-\sin\theta, \cos\theta)$ 이므로 다음이 성립한다.

$$\begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix}. \tag{6.16}$$

즉,

$$b = -\sin\theta, \qquad d = \cos\theta$$

이다.

따라서, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$에서, $\mathbf{x} = (x_1, x_2)^T$ 를 반시계방향으로 θ 만큼 회전시켜 얻은 $\mathbf{y} = (y_1, y_2)^T$ 는 아래와 같다.

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A\mathbf{x}. \tag{6.17}$$

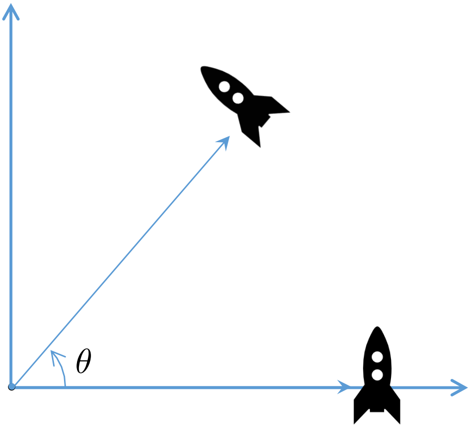

Figure 6.3: 회전변환에 의해 변환된 이미지

Example

반시계방향으로 90도 회전시키는 선형변환을 이루는 행렬이 아래와 같음을 보여라.

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{6.18}$$

6.3.2 길이변환

다음 선형변환은 수평방향과 수직방향으로 k 배 시켜주는 변환이다.

$$\begin{aligned}\mathbf{y} &= A\mathbf{x} \\ \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} &= \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.\end{aligned} \tag{6.19}$$

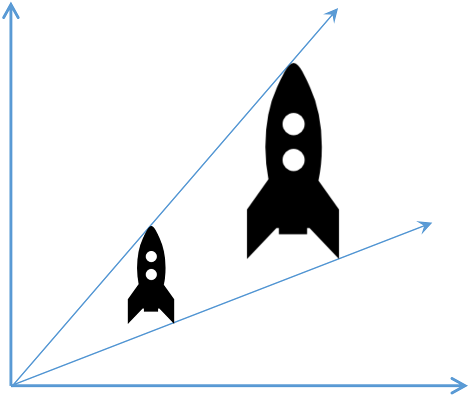

Figure 6.4: 길이변환에 의해 크기가 2배로 확대된 이미지

6.3.3 내분점의 선형변환

Definition

두 벡터 **a**, **b** 를 $m : n$ 으로 내분하는 벡터 **c** 는

$$\mathbf{c} = \frac{n\mathbf{a} + m\mathbf{b}}{m + n} \tag{6.20}$$

로 구해지며, 그 모습은 Figure 6.5와 같다.

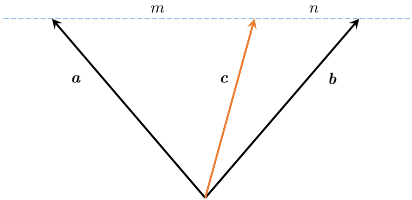

Figure 6.5: 벡터 **a**, **b** 를 $m : n$ 으로 내분하는 벡터 **c**

Theorem (정리 6.2)

두 벡터 $\mathbf{a}, \mathbf{b}$ *가 선형변환* $\mathbf{y} = A\mathbf{x}$ *에 의해* $\mathbf{a}' = A\mathbf{a}$, $\mathbf{b}' = A\mathbf{b}$ *로 변환되었으면,* $\mathbf{a}, \mathbf{b}$ *를* $m : n$ *으로 내분하는* $\mathbf{c}$ *에 대해서도*

$$\mathbf{c}' = A\mathbf{c} = \frac{n\mathbf{a}' + m\mathbf{b}'}{m+n} \tag{6.21}$$

이 성립한다.

Proof.

$$A\mathbf{c} = A\left(\frac{n\mathbf{a} + m\mathbf{b}}{m+n}\right) = \left(\frac{n}{m+n}\right)A\mathbf{a} + \left(\frac{m}{m+n}\right)A\mathbf{b} = \frac{n\mathbf{a}' + m\mathbf{b}'}{m+n}$$

□

6.3.4 직교행렬과 직교변환

Definition

정방행렬 $A_{n\times n}$ 이

$$A^T A = AA^T = I_n \tag{6.22}$$

을 만족하면 A 는 직교행렬(orthogonal matrix)이라고 한다.

Proposition (6.3)

A 가 직교행렬이면 $A^{-1} = A^T$ 이다.

Example

직교행렬의 예:

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad B = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Theorem (정리 6.4)

정방행렬 $A_{n\times n}$이 직교행렬이 되기 위한 필요충분조건은 아래와 같다.

(1) $\mathbf{a}_i^T\mathbf{a}_j = \delta_{ij}, \quad i,j = 1,2,\ldots,n$

(2) $\tilde{\mathbf{a}}_i^T\tilde{\mathbf{a}}_j = \delta_{ij}, \quad i,j = 1,2,\ldots,n.$

여기서 δ_{ij}의 값은 $i=j$ 이면 1, $i\neq j$ 이면 0 이다.

Proof.

$A = (\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \ldots, \tilde{\mathbf{a}}_n)$ 이라 표현하면,

$$A^TA = \begin{pmatrix} \tilde{\mathbf{a}}_1^T \\ \tilde{\mathbf{a}}_2^T \\ \vdots \\ \tilde{\mathbf{a}}_n^T \end{pmatrix} (\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \ldots, \tilde{\mathbf{a}}_n) = \begin{pmatrix} \tilde{\mathbf{a}}_1^T\tilde{\mathbf{a}}_1 & \tilde{\mathbf{a}}_1^T\tilde{\mathbf{a}}_2 & \cdots & \tilde{\mathbf{a}}_1^T\tilde{\mathbf{a}}_n \\ \tilde{\mathbf{a}}_2^T\tilde{\mathbf{a}}_1 & \tilde{\mathbf{a}}_2^T\tilde{\mathbf{a}}_2 & \cdots & \tilde{\mathbf{a}}_2^T\tilde{\mathbf{a}}_n \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{a}}_n^T\tilde{\mathbf{a}}_1 & \tilde{\mathbf{a}}_n^T\tilde{\mathbf{a}}_2 & \cdots & \tilde{\mathbf{a}}_n^T\tilde{\mathbf{a}}_n \end{pmatrix}. \tag{6.23}$$

Proof (continue).

A 가 직교행렬이면 $A^TA = I_n$ 이므로

$$\begin{pmatrix} \tilde{\mathbf{a}}_1^T\tilde{\mathbf{a}}_1 & \tilde{\mathbf{a}}_1^T\tilde{\mathbf{a}}_2 & \cdots & \tilde{\mathbf{a}}_1^T\tilde{\mathbf{a}}_n \\ \tilde{\mathbf{a}}_2^T\tilde{\mathbf{a}}_1 & \tilde{\mathbf{a}}_2^T\tilde{\mathbf{a}}_2 & \cdots & \tilde{\mathbf{a}}_2^T\tilde{\mathbf{a}}_n \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{a}}_n^T\tilde{\mathbf{a}}_1 & \tilde{\mathbf{a}}_n^T\tilde{\mathbf{a}}_2 & \cdots & \tilde{\mathbf{a}}_n^T\tilde{\mathbf{a}}_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{6.24}$$

따라서 $\tilde{\mathbf{a}}_i^T\tilde{\mathbf{a}}_j = \delta_{ij}$ 가 성립한다.

이제

$$A = \begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_n^T \end{pmatrix}$$

라 표현하자. 그러면

Proof (continue).

$$AA^T = \begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_n^T \end{pmatrix} (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = \begin{pmatrix} \mathbf{a}_1^T\mathbf{a}_1 & \mathbf{a}_1^T\mathbf{a}_2 & \cdots & \mathbf{a}_1^T\mathbf{a}_n \\ \mathbf{a}_2^T\mathbf{a}_1 & \mathbf{a}_2^T\mathbf{a}_2 & \cdots & \mathbf{a}_2^T\mathbf{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_n^T\mathbf{a}_1 & \mathbf{a}_n^T\mathbf{a}_2 & \cdots & \mathbf{a}_n^T\mathbf{a}_n \end{pmatrix} \tag{6.25}$$

이고 $AA^T = I_n$ 이어야 하므로, $\mathbf{a}_i^T\mathbf{a}_j = \delta_{ij}$ 가 성립한다. □

Remark 행렬 A 가 직교행렬이면 열벡터 $\tilde{\mathbf{a}}_i$ 들은 직교정규벡터(orthonormal vector)이다.

Remark 행렬 A 가 직교행렬이면 행벡터 $\mathbf{a}_i$ 들은 직교정규벡터(orthonormal vector)이다.

Definition

선형변환 $\mathbf{y} = A\mathbf{x}$ 에서 변환행렬 $A_{n\times n}$이 직교행렬이면, 선형변환 $\mathbf{y} = A\mathbf{x}$ 를 직교변환 (orthogonal transformation)이라고 한다.

Theorem (정리 6.5)

정방행렬 $A_{n\times n}$가 직교행렬이고, $\mathbf{y} = A\mathbf{x}$ 를 만족할 때,

$$\|\mathbf{x}\| = \|\mathbf{y}\| \tag{6.26}$$

를 만족한다.

Proof.

$$\|\mathbf{y}\| = \|A\mathbf{x}\| = \sqrt{(A\mathbf{x})^T(A\mathbf{x})} = \sqrt{\mathbf{x}^T A^T A\mathbf{x}} = \sqrt{\mathbf{x}^T\mathbf{x}} = \|\mathbf{x}\|.$$

□

Theorem (정리 6.6)

직교변환에 의해 $\mathbf{y}_1 = A\mathbf{x}_1$, $\mathbf{y}_2 = A\mathbf{x}_2$ *가 성립할 때,*

$$\mathbf{y}_1^T\mathbf{y}_2 = \mathbf{x}_1^T\mathbf{x}_2 \tag{6.27}$$

가 성립한다. 즉 벡터의 내적값이 변환에 대해 불변이다.

Proof.

$$\mathbf{y}_1^T\mathbf{y}_2 = (A\mathbf{x}_1)^T(A\mathbf{x}_2) = \mathbf{x}_1^T A^T A\mathbf{x}_2 = \mathbf{x}_1^T\mathbf{x}_2.$$

□

Example

회전변환 $\mathbf{y} = A\mathbf{x}$ 는 직교변환이다. $\mathbb{R}^2$ 상의 회전변환에 대응하는 행렬
$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ 는 $A^TA = AA^T = I_2$ 이다.

6.4 아핀변환(Affine Transformation)

직선함수 $f(x) = 2x$ 가 선형함수 임에도 절편이 있는 직선함수

$$f(x) = 1 + 2x \tag{6.28}$$

는 선형함수의 정의를 만족하지 않는다. 이러한 직선함수를 아핀함수(affine function)라 한다.

Definition

함수가 $f(x) = a + bx$ 의 형태 (a, b 는 상수) 일 때, $f(x)$ 를 아핀함수(affine function) 이라 한다.

Definition

벡터 $\mathbf{x}$ 에 대해 변환

$$\mathbf{y} = A\mathbf{x} + \mathbf{b} \tag{6.29}$$

를 아핀변환(affine transformation)이라고 한다.

Remark 아핀변환은 선형변환 $A\mathbf{x}$ 을 취한 후, 다시 $\mathbf{b}$ 방향으로 평행이동한 변환이다.

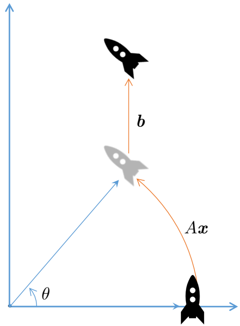

Figure 6.6: 아핀변환 $\mathbf{y} = A\mathbf{x} + \mathbf{b}$ 의 모습

Remark 선형변환은 $\mathbf{b} = \mathbf{0}$ 인 아핀변환이다.