행렬대수 1 - 사영과 QR 분해

시립대학교 통계학과

2019년 6월 3일

1 두 벡터의 사영

선형독립인 두 벡터 a_1 과 a_2 가 있다고 하자. 벡터 a_1 과 같은 방향을 가지면서 벡터 a_2 에 가장 가까운 벡터를 $proj_{a_1}(a_2)$ 라고 정의하고 이를 벡터 a_1 방향으로 벡터 a_2 의 사영(projection) 이라고 부른다.

그러면 이러한 사영은 어떻게 구할 수 있나? 벡터 a_2 의 사영은 벡터 a_1 방향에 있으므로 어떤 실수 c 가 있어서 다음과 같이 표시할 수 있다.

$$proj_{a_1}(a_2) = ca_1$$

이제 사영 ca_1 과 벡터 a_2 의 거리 d(c) 를 생각하면 다음과 같다.

$$d^{2}(c) = \|a_{2} - ca_{1}\|^{2}$$

$$= (a_{2} - ca_{1})^{t}(a_{2} - ca_{1})$$

$$= a_{2}^{t}a_{2}^{t} - 2ca_{2}^{t}a_{1} + c^{2}a_{1}^{t}a_{1}^{t}$$

위에서 ||a|| 는 벡터 a의 길이를 나타낸다.

$$d(a) = ||a|| = \sqrt{a^t a}$$

상수 c 는 거리 d(c)를 최소로 만드는 수이다. $d^2(c)$ 은 c 에 대하여 미분 가능한 2차 함수이며 아래로 볼록한 함수이므로 이를 미분하여 c 를 구할 수 있다.

$$\frac{\partial d^2(c)}{\partial c} = -2a_2^t a_1 + 2ca_1^t a_1^t = 0$$

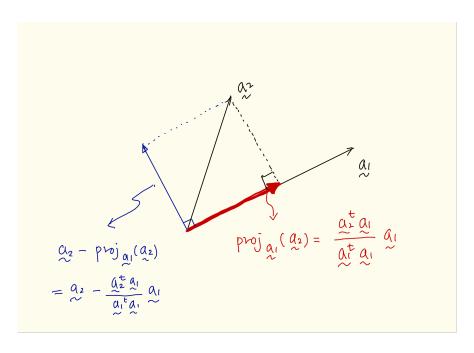


Figure 1: "벡터의 사영

위의 방적식으로 부터 c를 얻고

$$c = \frac{a_2^t a_1}{a_1^t a_1^t}$$

다음과 같이 벡터 a_1 방향으로 벡터 a_2 의 사영을 나타낼 수 있다.

$$proj_{a_1}(a_2) = \frac{a_2^t a_1}{a_1^t a_1} a_1 \tag{1}$$

이제 위의 두 벡터의 사영을 이용하면 벡터 a_1 과 직교하는 벡터 \tilde{q}_2 를 다음과 같이 찾을 수 있다.

$$\tilde{q}_2 = a_2 - proj_{a_1}(a_2) = a_2 - \frac{a_2^t a_1}{a_1^t a_1} a_1$$

두 벡터 a_1 와 \tilde{q}_2 의 직교성은 다음과 같이 보일 수 있다.

$$a_{1}^{t}\tilde{q}_{2} = a_{1}^{t} \left(a_{2} - \frac{a_{2}^{t}a_{1}}{a_{1}^{t}a_{1}} a_{1} \right)$$

$$= a_{1}^{t} a_{2} - \frac{a_{2}^{t}a_{1}}{a_{1}^{t}a_{1}} a_{1}^{t} a_{1}$$

$$= a_{1}^{t} a_{2} - a_{2}^{t} a_{1}$$

$$= 0$$

$$(2)$$

이제 두 벡터 q_1 과 q_2 를 다음과 같이 정규직교벡터로 만들 수 있다.

$$q_1 = a_1 / \|a_1\|$$

 $q_2 = \tilde{q}_2 / \|\tilde{q}_2\|$

1.1 Gram-Schmidt 방법

서로 독립인 n차원의 벡터들이 p개 있을때

$$a_1, a_2, \ldots, a_p$$

이들이 만드는 열공간을 C라고 하자.

$$C = span\{a_1, a_2, \dots, a_p\}$$

$$= \{ c_1 a_1 + c_2 a_2 + \dots + c_p a_p \mid \text{ all possible real values of } c_1, c_2, \dots, c_p \}$$
(3)

이제 우리는 위와 동일한 열공간 C 만드는 정규직교 벡터들을 찾는 방법을 알아보고자한다.

$$q_1, q_2, ..., q_p$$
 where $q_i^t q_i = 0$, $q_i^t q_i = 1$

그리고

$$C = span\{q_1, q_2, \dots, q_p\} = span\{a_1, a_2, \dots, a_p\}$$
 (4)

이제 앞 절의 벡터의 사영에 대한 결과를 사용하여 다음과 같은 직교하는 p 개의 벡터들을 축차적으로 만들어 보자.

$$\tilde{q}_{1} = a_{1}
\tilde{q}_{2} = a_{2} - proj_{\tilde{q}_{1}}(a_{2})
\tilde{q}_{3} = a_{3} - proj_{\tilde{q}_{1}}(a_{3}) - proj_{\tilde{q}_{2}}(a_{3})
\tilde{q}_{4} = a_{4} - proj_{\tilde{q}_{1}}(a_{4}) - proj_{\tilde{q}_{2}}(a_{4}) - proj_{\tilde{q}_{3}}(a_{4})
...
\tilde{q}_{p} = a_{p} - \sum_{k=1}^{p} proj_{\tilde{q}_{k}}(a_{p})$$
(5)

축차적으로 만든 벡터들을 정규벡터로 만들면 원래의 벡터들 a_1, a_2, \ldots, a_p 이 생성하는 동일한 열공간을 만드는 정규직교 벡터 q_1, q_2, \ldots, q_p 를 만들 수 있다.

$$q_i = \tilde{q}_i / \|\tilde{q}_i\|, \quad i = 1, 2, \dots, p \tag{6}$$

Gram—Schmidt 방법으로 만든 벡터들이 직교하는 것은 다음과 같이 증명할 수 있다. 먼저 식 (2) 에 의하여 \tilde{q}_1 과 \tilde{q}_2 는 직교한다. 이제 임의의 i에 대하여 $\tilde{q}_1, \tilde{q}_2, \cdots, \tilde{q}_{i-1}$ 벡터들이 직교한다고 가정하자. 모든 $1 \leq j \leq i-1$ 에 대하여

$$\begin{split} \tilde{\mathbf{q}}_{j}^{t} \tilde{\mathbf{q}}_{i} &= \tilde{\mathbf{q}}_{j}^{t} \left[\mathbf{a}_{i} - \sum_{k=1}^{i-1} proj_{\tilde{\mathbf{q}}_{k}}(\mathbf{a}_{i}) \right] \\ &= \tilde{\mathbf{q}}_{j}^{t} \left[\mathbf{a}_{i} - proj_{\tilde{\mathbf{q}}_{j}}(\mathbf{a}_{i}) \right] - \left[\sum_{\substack{1 \leq k \leq i-1 \\ k \neq j}} \tilde{\mathbf{q}}_{j}^{t} \ proj_{\tilde{\mathbf{q}}_{k}}(\mathbf{a}_{i}) \right] \\ &= 0 + 0 \end{split}$$

위에서 마지막 단계의 직교성은 다음과 같은 사실로 부터 유도된다.

- $1. \ a_i proj_{\tilde{q}_i}(a_i)$ 는 \tilde{q}_i^t 와 직교한다.
- 2. 가정에 의하여 $ilde{q}_1, ilde{q}_2, \cdots, ilde{q}_{i-1}$ 는 직교하고 $proj_{ ilde{q}_k}(a_i)$ 는 $ilde{q}_k$ 와 같은 방향을 가진다.

$$\tilde{q}_{j}^{t} proj_{\tilde{q}_{k}}(a_{i}) = 0$$
 for $1 \leq j, k \leq i - 1, k \neq j$

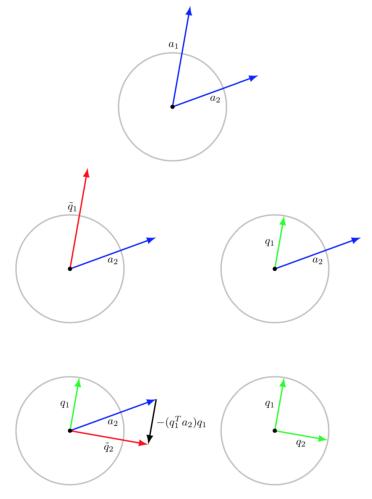


Figure 5.3 Gram—Schmidt algorithm applied to two 2-vectors a_1 , a_2 . Top. The original vectors a_1 and a_2 . The gray circle shows the points with norm one. Middle left. The orthogonalization step in the first iteration yields $\tilde{q}_1 = a_1$. Middle right. The normalization step in the first iteration scales \tilde{q}_1 to have norm one, which yields q_1 . Bottom left. The orthogonalization step in the second iteration subtracts a multiple of q_1 to yield the vector \tilde{q}_2 , which is orthogonal to q_1 . Bottom right. The normalization step in the second iteration scales \tilde{q}_2 to have norm one, which yields q_2 .

Figure 2: Gram-Schmidt 방법을 설명한 그림(출처:Introduction to Applied Linear Algebra by Boyd and Vandenberghe, 2019)

식 (5)과 (6)의 알고리즘을 Gram-Schmidt 방법이라고 부른다. 위의 두 식에 의한 알고리즘을 다음과 같은 사실을 이용하면 좀 더 간단한 방법의 알고리즘이 나온다.

$$proj_{ ilde{q}_k}(a_l) = rac{a_l^t ilde{q}_k}{ ilde{q}_k^t ilde{q}_k} ilde{q}_k = rac{a_l^t ilde{q}_k}{\| ilde{q}_k\|^2} ilde{q}_k = (a_l^t q_k) q_k$$

- p개의 벡터 $a_1, a_2, ..., a_p$ 에 대하여
- for i = 1, 2, ..., p

$$1. \ ilde{q}_i = a_i - (q_1^t a_i) q_1 - \dots - (q_{i-1}^t a_i) q_{i-1}$$
 (직교화)

$$2. \ \mathbf{q}_i = \tilde{\mathbf{q}}_i / \|\mathbf{q}_i\| \ (정규화)$$

2 최소제곱법과 사영

회귀계수벡터의 값을 구하는 최소제곱법의 기준을 다시 살펴보자.

$$\min_{\beta} \sum_{i=1}^{n} (y_i - x_i^t \beta)^2 = \min_{\beta} (y - X\beta)^t (y - X\beta) = \min_{\beta} ||y - X\beta||^2$$
 (7)

위에서 Xeta는 행렬 X의 열벡터 a_0,a_1,\ldots,a_p 로 이루어진 선형조합이다.

$$X\beta = [a_0 \ a_1 \ \cdots a_p]\beta = \beta_0 a_0 + \beta_1 a_1 + \cdots + \beta_p a_p$$

따라서 최소제곱법으로 구한 회귀계수 벡터 $\hat{m{\beta}}$ 는 반응값 벡터 $m{y}$ 와 $m{X}m{eta}$ 의 거리가 최소가 되도록 만들어 준다.

$$\hat{oldsymbol{eta}} = (X^t X)^{-1} X^t y, \quad \min_{oldsymbol{eta}} \| oldsymbol{y} - X oldsymbol{eta} \|^2 = \left\| oldsymbol{y} - X \hat{oldsymbol{eta}}
ight\|^2$$

따라서 예측값 벡터 \hat{y} 는 행렬 X의 열벡터로 생성한 열공간 방향으로 반응값 벡터 y를 사영한 벡터이다.

$$\hat{\mathbf{y}} = X\hat{\mathbf{\beta}} = X(X^tX)^{-1}X^t\mathbf{y}$$

위에서 행렬 $P=X(X^tX)^{-1}X^t$ 를 열공간 C(X)의 사영행렬(projection matirx)라고 부른다.

3 QR 분해

식 (5)과 (6)에 주어진 Gram—Schmidt 방법을 원래 벡터들 a_1, a_2, \ldots, a_p 에 대하여 다시 다음과 같이 나타낼 수 있다.

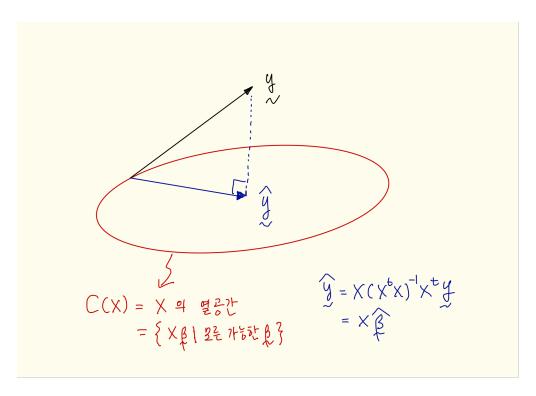


Figure 3: 최소제곱법을 설명한 그림

$$a_{1} = \tilde{q}_{1}$$

$$= \|\tilde{q}_{1}\| q_{1}$$

$$a_{2} = \tilde{q}_{2} + proj_{\tilde{q}_{1}}(a_{2})$$

$$= \tilde{q}_{2} + \frac{a_{2}^{t}\tilde{q}_{1}}{\tilde{q}_{1}^{t}\tilde{q}_{1}}\tilde{q}_{1}$$

$$= (a_{2}^{t}q_{1})q_{1} + \|\tilde{q}_{2}\| q_{2}$$

$$a_{3} = \tilde{q}_{3} + proj_{\tilde{q}_{1}}(a_{3}) + proj_{\tilde{q}_{2}}(a_{3})$$

$$= \tilde{q}_{3} + \frac{a_{3}^{t}\tilde{q}_{1}}{\tilde{q}_{1}^{t}\tilde{q}_{1}}\tilde{q}_{1} + \frac{a_{3}^{t}\tilde{q}_{2}}{\tilde{q}_{2}^{t}\tilde{q}_{2}}\tilde{q}_{2}$$

$$= (a_{3}^{t}q_{1})q_{1} + (a_{3}^{t}q_{2})q_{2} + \|\tilde{q}_{3}\| q_{3}$$

$$\cdots$$

$$a_{p} = (a_{p}^{t}q_{1})q_{1} + (a_{p}^{t}q_{2})q_{2} + \cdots + (a_{p}^{t}q_{p-1})q_{p-1} + \|\tilde{q}_{p}\| q_{p}$$

즉 위의 축차식을 보면 원래 벡터 a_i 는 Gram-Schmidt 방법으로 구한 정규직교벡터 q_1, q_2, \ldots, q_p 의 선형 조합으로 나타낼 수 있다.

이제 Gram—Schmidt 방법으로 구한 정규직교벡터들 q_1,q_2,\ldots,q_p 을 모아놓은 행렬을 Q라고 하고 위에서 a_i 들이 직교행렬의 선형조합으로 표시될때 계수들을 모아놓는 상삼각행렬을 R 이라고 하자. 그려면 다음과 같은 QR 분해가 주어진다.

$$A = QR \tag{8}$$

여기서

$$egin{aligned} m{Q} &= [m{q}_1 \ m{q}_2 \ \dots \ m{q}_p], \quad m{Q}^t m{Q} = m{I} \ &= egin{bmatrix} \|m{ ilde{q}}_1\| & m{a}_2^t m{q}_1 & m{a}_3^t m{q}_1 & \dots & m{a}_p^t m{q}_1 \ 0 & \|m{ ilde{q}}_2\| & m{a}_3^t m{q}_2 & \dots & m{a}_p^t m{q}_2 \ 0 & 0 & \|m{ ilde{q}}_3\| & \dots & m{a}_p^t m{q}_3 \ & & \dots & & \ 0 & 0 & 0 & \dots & \|m{ ilde{q}}_p\| \end{bmatrix} \end{aligned}$$

4 최소제곱법과 QR 분해

다시 식 (7)의 최소제곱법을 고려하자. 실제로 $\hat{\beta}$ 을 구하는 경우 많은 통계계산 프로그램에서 실제로 역행렬 $(X^tX)^{-1}$ 을 계산하는 경우는 드물다. 실제 최소제곱법을 푸는 알고리즘으로 많은 경우 QR 분해법을 사용한다.

 $n \times (p+1)$ 행렬 X의 QR 분해가 다음과 같이 주어졌다고 하자. 또한 r(X) = p+1 < n이라고 가정하자.

$$X = Q_1 R_1$$

그러면 행렬 X는 다음과 같은 확장된 분해를 가진다.

$$X = QR = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

위에서 Q_2 는 행렬 Q_1 의 열들과 직교하는 n-(p+1) 개의 추가 정규직교 벡터들로 이루어진 행렬이다. Q_1 과 Q_2 는 각각 $n \times (p+1)$, $n \times (n-p-1)$ 의 행렬이다. 따라서 행렬 $Q=[Q_1 \ Q_2]$ 는 $n \times n$ 직교행렬이다 ($Q^tQ=QQ^t=I$). 여기서 R_1 는 $(p+1) \times (p+1)$ 상삼각행렬(upper diagonal matrix)고 0은 차원이 $(n-p-1) \times (p+1)$ 인 영행렬이다.

이제 $Q^tQ = I$ 를 이용하여 잔차제곱합 $\|y - X\beta\|^2$ 를 다음과 같이 분해할 수 있다.

$$||y - X\beta||^{2} = (y - X\beta)^{t}(y - X\beta)$$

$$= (y - X\beta)^{t}QQ^{t}(y - X\beta)$$

$$= (Q^{t}y - Q^{t}X\beta)^{t}(Q^{t}y - Q^{t}X\beta)$$

$$= (c - R\beta)^{t}(c - R\beta)$$

$$= \left(\begin{bmatrix} c_{1} \\ c_{2} \end{bmatrix} - \begin{bmatrix} R_{1} \\ 0 \end{bmatrix} \beta \right)^{t} \left(\begin{bmatrix} c_{1} \\ c_{2} \end{bmatrix} - \begin{bmatrix} R_{1} \\ 0 \end{bmatrix} \beta \right)$$

$$= (c_{1} - R_{1}\beta)^{t}(c_{1} - R_{1}\beta) + c_{2}^{t}c_{2}$$

$$= ||c_{1} - R_{1}\beta||^{2} + ||c_{2}||^{2}$$
(9)

여기서

$$c = oldsymbol{Q}^t y = egin{bmatrix} oldsymbol{Q}_1^t \ oldsymbol{Q}_2^t \end{bmatrix} y = egin{bmatrix} oldsymbol{Q}_1^t y \ oldsymbol{Q}_2^t y \end{bmatrix} = egin{bmatrix} oldsymbol{c}_1 \ oldsymbol{c}_2 \end{bmatrix}$$

위의 식 (9)를 보면 벡터 $c_2=Q_2^ty$ 는 β 와 관계가 없으므로 잔차제곱합 $\|y-X\beta\|^2$ 을 최소화하는 β 는 $\|c_1-R_1\beta\|^2$ 을 0으로 만드는 것이다. 즉 $R_1\beta=c_1$ 를 만족하는 β 가 최소제곱 추정량이다.

$$\min_{\beta} \|y - X\beta\|^2 = \min_{\beta} \|c_1 - R_1\beta\|^2 + \|c_2\|^2$$

X가 완전계수 행렬이므로 상삼각행렬인 R_1 도 완전계수이다. 따라서 방정식 $R_1\beta=c_1$ 는 유일한 해는 상삼각행렬의 성질을 이용하여 축차식으로 쉽게 구할 수 있다.

$$\hat{\beta} = R_1^{-1}c_1$$

위의 추정량은 실제 역행렬을 구하지 않고 QR 분해를 이용하여 회귀계수를 구하는 방법 중에 하나이다.

5 예제

5.1 Gram-Schmidt 방법

아래와 같이 4차원 벡터 3개가 있다.

$$a_{1} = \begin{bmatrix} -1\\1\\-1\\1 \end{bmatrix} \qquad a_{2} = \begin{bmatrix} -1\\3\\-1\\3 \end{bmatrix} \qquad a_{3} = \begin{bmatrix} 1\\3\\5\\7 \end{bmatrix}$$
 (10)

위의 벡터 a_1, a_2, a_3 에 대하여 Gram-Schmidt 방법을 적용해보자.

ullet i=1. 먼저 $\| ilde{q}_1\|=\|a_1\|=2$ 이므로 첫번째 벡터 q_1 를 만든다.

$$m{q}_1 = ilde{m{q}}_1 / \| ilde{m{q}}_1 \| = egin{bmatrix} -1/2 \ 1/2 \ -1/2 \ 1/2 \end{bmatrix}$$

ullet i=2. 이제 두번째 직교벡터 $oldsymbol{q}_2$ 를 만들자. $oldsymbol{q}_1^toldsymbol{a}_2=4$ 이므로

$$ilde{q_2} = a_2 - (q_1^t a_2) q_1 = egin{bmatrix} -1 \ 3 \ -1 \ 3 \end{bmatrix} - 4 egin{bmatrix} -1/2 \ 1/2 \ -1/2 \ 1/2 \end{bmatrix} = egin{bmatrix} 1 \ 1 \ 1 \ 1 \end{bmatrix}$$

그리고 $\| ilde{q}_2\|=2$ 이므로

$$q_2 = \tilde{q}_2 / \|\tilde{q}_2\| = egin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$$

• i = 3 마지막으로 $q_1^t a_3 = 2$, $q_2^t a_3 = 8$ 이므로

$$\tilde{q}_{3} = a_{3} - (q_{1}^{t}a_{3})q_{1} - (q_{2}^{t}a_{3})q_{2} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \end{bmatrix} - 2 \begin{bmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix} - 8 \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ 2 \\ 1/2 \end{bmatrix}$$

또한 $\|\tilde{q}_3\|=4$ 이므로

$$q_3 = \tilde{q}_3 / \|\tilde{q}_3\| = egin{bmatrix} -1/2 \\ -1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$$

따라서 Gram-Schmidt 방법으로 만든 정규직교벡터는 다음과 같다.

$$q_1 = \begin{bmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix} \quad q_2 = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} \quad q_3 = \begin{bmatrix} -1/2 \\ -1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$$

6 QR 분해

식 (10) 에서 주어진 백터들을 열로 가진 행렬 A의 QR분해를 구해보자.

$$A = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 3 & 3 \\ -1 & -1 & 5 \\ 1 & 3 & 7 \end{bmatrix}$$

앞의 예제에서 구한 직교벡터를 그대로 이용하면 Q는 쉽게 구해진다.

$$Q = \begin{bmatrix} -1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & -1/2 \\ -1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 \end{bmatrix}$$

또한 식 (8)에 주어진 공식을 이용하면 행렬 R은 다음과 같이 구할 수 있다.

$$R = \begin{bmatrix} \|\tilde{q}_1\| & a_2^t q_1 & a_3^t q_1 \\ 0 & \|\tilde{q}_2\| & a_3^t q_2 \\ 0 & 0 & \|\tilde{q}_3\| \end{bmatrix} = \begin{bmatrix} 2 & 4 & 2 \\ 0 & 2 & 8 \\ 0 & 0 & 4 \end{bmatrix}$$