Transformation

Distance, Variance,
&
Decomposition

H. Park

HUFS

유클리디안 거리(Euclidian Distance)

- x_i^T = i-번째 개체벡터 , x_j^T = j-번째 개체벡터

$$X = \begin{pmatrix} 165 & 63 & 85 \\ 170 & 70 & 90 \\ 180 & 75 & 105 \\ 173 & 70 & 85 \end{pmatrix} \rightarrow$$

❧ Euclidian Distance = ______________________거리

$$\begin{aligned} d_{13} &= \|\boldsymbol{x}_1 - \boldsymbol{x}_3\| = \sqrt{(\boldsymbol{x}_1 - \boldsymbol{x}_3)^T (\boldsymbol{x}_1 - \boldsymbol{x}_3)} \\ &= \sqrt{\sum_{k=1}^{3} (x_{1k} - x_{3k})^2} \end{aligned}$$

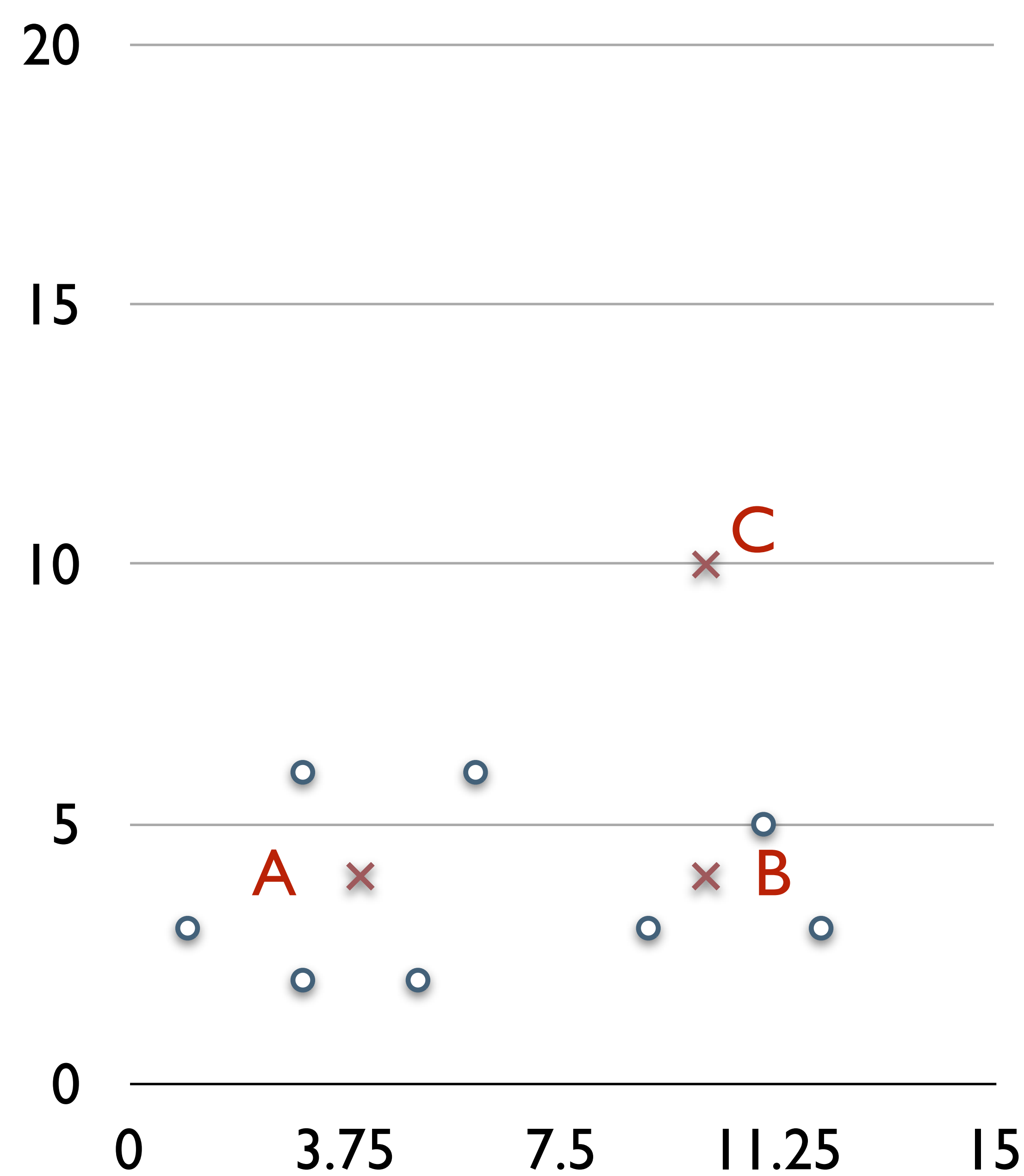
(Q) 4번째 개체는 1번째와 3번째 중 누구와 거리가 가까운지 유클리디안 거리를 이용해 답해 보아라.

마할라노비스 거리(Mahalanobis Distance)

- 유클리디안 거리 : AB BC
- 통계학적 거리 : AB BC

$$sd_{AB} = \sqrt{\frac{(x_{A1} - x_{B1})^2}{s_{11}} + \frac{(x_{A2} - x_{B2})^2}{s_{22}}}$$

$$sd_{BC} = \sqrt{\frac{(x_{B1} - x_{C1})^2}{s_{11}} + \frac{(x_{B2} - x_{C2})^2}{s_{22}}}$$

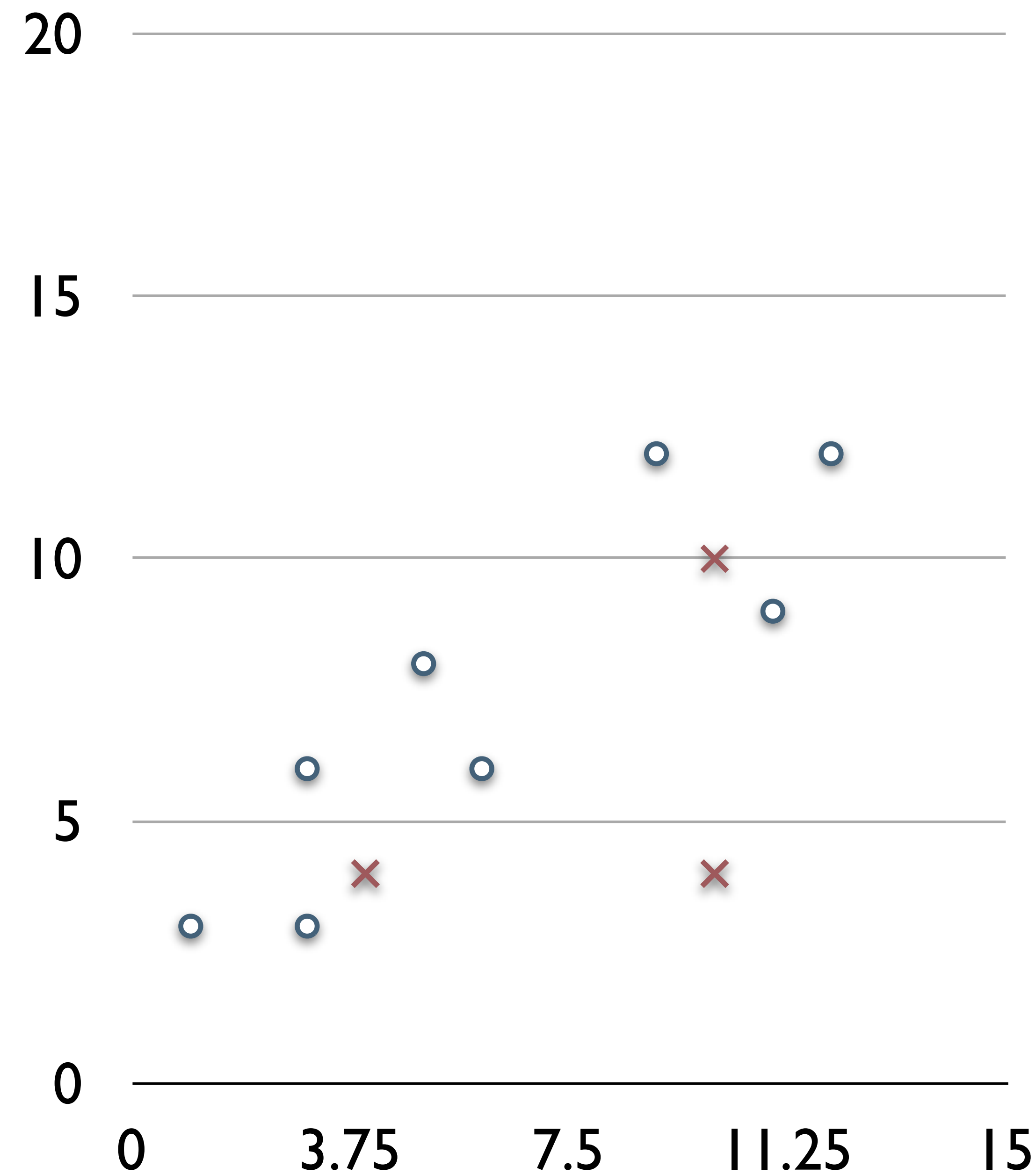


What if ....

$Cov(x_1, x_2) = s_{12} > 0$

= 마할라노비스 거리

(Mahalanobis Distance)



(Q) 4번째 개체는 1번째와 3번째 중 누구와 거리가 가까운지 마할라노비스 거리를 이용해 답해 보아라(s11= 16.2 , s22=12.5, s12=11.0)

sub	x1	x2
1	1	3
2	3	6
3	3	3
4	6	6
5	5	8
6	9	12
7	11	9
8	12	12

$$S = \begin{pmatrix} 16.2 & 11.0 \\ 11.0 & 12.5 \end{pmatrix}$$

$$S^{-1} = \begin{pmatrix} 0.15 & -0.13 \\ -0.13 & 0.19 \end{pmatrix}$$

고유값 & 고유벡터
(eigenvalue, eigenvector)

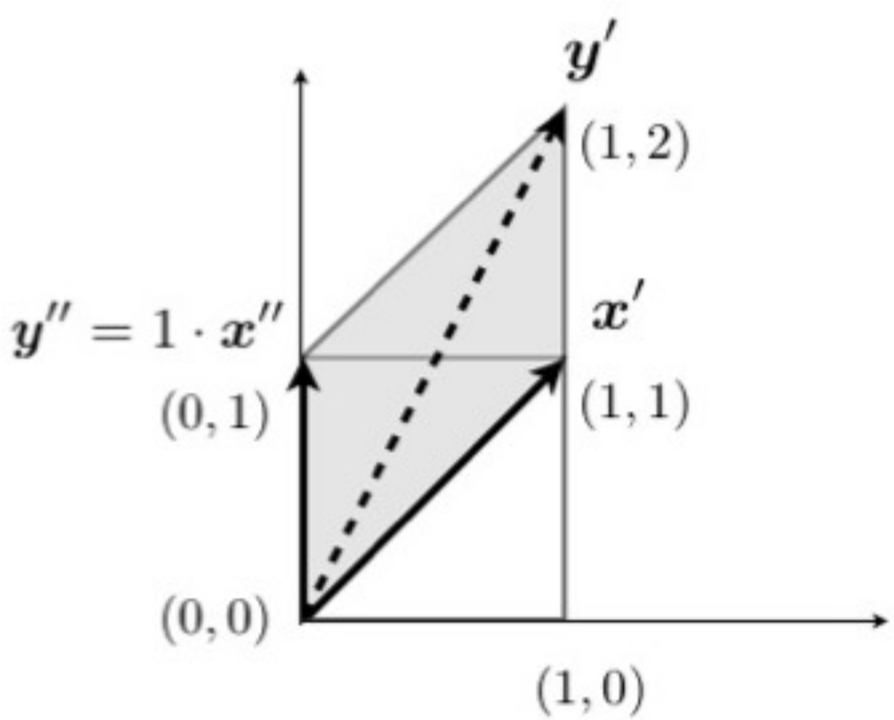
Definition 정방행렬 A에 대하여, $A\boldsymbol{x} = \lambda\boldsymbol{x}$가 성립할 때, λ를 고유값(eigenvalue), $\boldsymbol{x}$를 λ에 따른 고유벡터(eigenvector)라고 한다. 이 때, $\boldsymbol{x}$는 영이 아닌 벡터이다.

Example 아래 행렬 A은 $\Re^2 \longrightarrow \Re^2$로 변환시키는 선형변환 $\boldsymbol{y} = A\boldsymbol{x}$에 사용된 행렬이며 아래와 같다.

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{12.2}$$

위 선형변환은 그림 12.1와 같이 정사각형을 평행사변형으로 변형시키는 역할을 한다.

[1]'eigen'이란 독일어로 '선천적', '고유한'이란 의미의 단어이다.



고유값 구하는 법

$Ax = \lambda x$

Question: 다음 행렬의 고유값을 특성화방정식을 이용하여 구하라.

$$A = \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix} \tag{12.15}$$

Answer: 특성화 방정식을 구하면 다음과 같다.

$$p(\lambda) = \det(A - \lambda I) = \det \begin{pmatrix} 2-\lambda & 2 \\ 1 & 3-\lambda \end{pmatrix} = 0$$

이는 $p(\lambda) = \lambda^2 - 5\lambda + 4 = 0$ 라는 특성화방정식이 성립되므로 $\lambda = 1,\ 4$가 된다.

따라서, 행렬 A의 고유값은 1과 4이다.

고유벡터 구하는 법

$$A\boldsymbol{x} = \lambda\boldsymbol{x}$$

Remark 행렬 A의 고유값이 λ일 때, 고유벡터는

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

을 만족하는 $\boldsymbol{x}$를 구함으로써 얻어진다.

Question: 아래 행렬의 고유값이 1임을 알 때, 해당되는 고유벡터를 구하라.

$$B = \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix} \tag{12.18}$$

Answer: 확장행렬을 구해서 기본행연산을 하면

$$(A - \lambda I \mid 0) = \left(\begin{array}{cc|c} 2-1 & 2 & 0 \\ 1 & 3-1 & 0 \end{array}\right) = \left(\begin{array}{cc|c} 1 & 2 & 0 \\ 1 & 2 & 0 \end{array}\right)$$

$$\xrightarrow{R_{2,1(-1)}} \left(\begin{array}{cc|c} 1 & 2 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

그러므로 $x_1 + 2x_2 = 0$을 만족하는 모든 (x_1, x_2)가 답이 된다. 따라서 고유벡터는

$$\boldsymbol{x} = \begin{pmatrix} -2t \\ t \end{pmatrix}, \quad t\text{는 0 아닌 실수} \tag{12.19}$$

Generalized Variance

$$S = \text{sample covariance matrix}$$ (positive definite matrix)

$$= \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{pmatrix}, \ \{\lambda_1, \ \lambda_2, \ \lambda_3\}$$ eigenvalues

(Q) How can we measure S with a single value?

$\longrightarrow$ Total Variation

$\longrightarrow$ Generalized Variance

Linear Transformation

- 선형변환 $A : \boldsymbol{x} \rightarrow \boldsymbol{y}$ ⟶ $\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{b}$

❧ 어파인변환 $\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{b}$, A^{-1} $\boldsymbol{exits}$

❧ 척도변환 $\boldsymbol{y} = A\boldsymbol{x}$, where A is a diagonal matrix

(Q) 벡터 x_1 , x_2 간의 마할로니비스 거리는 어파인변환을 한 y_1 , y_2 벡터 간의 마할라노비스 거리와 일치함을 보여라.

(Q) 벡터 x_1 , x_2 간의 상관행렬은 척도변환을 한 y_1 , y_2 벡터 간의 상관행렬과 일치함을 보여라.

$$R_x = D_x^{-\frac{1}{2}} S_x D_x^{-\frac{1}{2}}$$

$$\boldsymbol{y} = A\boldsymbol{x}, \quad A \;\text{ is a sym. matrix}$$

$$S_y = AS_xA^T = AS_xA$$

$$D_y = \mathrm{diag}(S_y) = \mathrm{diag}(AS_xA) = A\;\mathrm{diag}(S_x)A = AD_xA$$

$$D_y^{-\frac{1}{2}} = (AD_xA)^{-\frac{1}{2}} = A^{-\frac{1}{2}} D_x^{-\frac{1}{2}}$$

$$\begin{aligned} R_y = D_y^{-\frac{1}{2}} S_y D_y^{-\frac{1}{2}} \quad &= \quad A^{-\frac{1}{2}} D_x^{-\frac{1}{2}} (AS_xA) A^{-\frac{1}{2}} D_x^{-\frac{1}{2}} \\ &= \quad D_x^{-\frac{1}{2}} S_x D_x^{-\frac{1}{2}} \\ &= \quad R_x \end{aligned}$$

Spectral Decomposition
(스펙트럼 분해)

- $\Sigma_{n \times n}$ = Symmetric & Positive Definite matrix

$$\begin{aligned} \lambda_i &= \text{eigenvalues} \\ \boldsymbol{e}_i &= \text{eigenvectors} \end{aligned}$$

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$$

아래 행렬을 **R**을 이용해,스펙트럼 분해하고

$\Sigma = \Gamma \Lambda \Gamma^T$ 임을 확인하라

$$A = \begin{pmatrix} 2 & \sqrt{7} & 0 \\ \sqrt{7} & 8 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Singular Value Decomposition
(특이값 분해)

- for a matrix X_{nxp} , rank(X)=q (<= p)

$$\Sigma\Sigma^T \boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$$
$$\Sigma^T\Sigma \boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$$

→ eigenvectors & eigenvalues

$$\boldsymbol{u}_i^T\boldsymbol{u}_j = \delta_{ij}$$
$$\boldsymbol{v}_i^T\boldsymbol{v}_j = \delta_{ij}$$

→ U & V are orthogonal matrices

Example 아래 행렬 A를 특이값 분해(SVD)를 실시하라.

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

행렬 AA^T의 고유값과 고유벡터를 구하면

$$AA^T = \begin{pmatrix} 2 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix}$$

그러므로, $\lambda_1 = 5$, $\lambda_2 = 1$ 가 된다. 고유벡터를 구해보면,

$$\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{u}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

가 되어

$$U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

이다. 한편 A^TA의 고유값과 고유벡터를 구해보면

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

에서 고유값은 0, 1, 5가 된다. 이제 0 아닌 고유값은 AA^T때와 마찬가지로 1, 5가 되고 직교정규 고유벡터는

$$\boldsymbol{v}_1 = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ 0 \\ -\frac{1}{\sqrt{5}} \end{pmatrix}, \quad \boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

가 되어

$$V = \begin{pmatrix} \frac{2}{\sqrt{5}} & 0 \\ 0 & 1 \\ -\frac{1}{\sqrt{5}} & 0 \end{pmatrix}$$

$$\begin{aligned} A = \begin{pmatrix} 2 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{5} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{5}} & 0 & -\frac{1}{\sqrt{5}} \\ 0 & 1 & 0 \end{pmatrix} \\ &= UDV^T \end{aligned}$$

Cholesky Decomposition
(촐레스키 분해)

- Σ = Symmetric + Positive Definite matrix

(Q) 다음 행렬 X 에 대한 공분산 행렬을 S 라고 할 때,

$$X = \begin{pmatrix} 165 & 63 & 85 \\ 170 & 70 & 90 \\ 180 & 75 & 105 \\ 173 & 70 & 85 \end{pmatrix}$$

(1) S 의 e-value/e-vector 구해서 Spectral Decomposition 실시하라

(2) S 의 Cholesky Decomposition 실시하라

(3) X 의 S.V. Decomposition 실시하라

```r
data<-c(165, 63, 85, 170, 70, 90, 180, 75, 105, 173, 70, 85)
X<-matrix(data,4,3,byrow=T)
S<-cov(X)
U<-eigen(S)$vectors
D<-diag(eigen(S)$values)
U%*%D%*%t(U)
svd(S)
L<-t(chol(S))
L%*%t(L)
svd(X)
U<-svd(X)$u
V<-svd(X)$v
D<-diag(svd(X)$d)
```