선형대수학 in 통계

오정헌

1 | 6

내적과 공분산

$$x \cdot y = \|x\|\|y\|cos\theta$$

$$x \cdot y = \sum_{i=1}^{n} x_i y_i$$

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \Rightarrow Cov(x,y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}$$

$$cos\theta = \frac{x \cdot y}{\|x\|\|y\|} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} = Cor(X, Y)$$

2 | 6

직교 부분공간

$$SSTO = SSE + SSR$$

$$Y^T(I - H_0)Y = Y^T(I - H)Y + Y^T(H - H_0)Y$$

$$\Rightarrow (I - H)(H - H_0) = 0$$

$$\underbrace{(I - H)}_{A} \perp \underbrace{(H - H_0)}_{B}$$

$$\begin{cases} Col(A) = Row(A) \\ Row(A) \perp Null(A) \\ Col(A) \perp Col(B) \end{cases} \Rightarrow \begin{array}{c} Null(A) = Col(B) \\ Null(I - H) = Col(H - H_0) \end{array}$$

3 | 6

회귀 계수와 Gram Schmidt

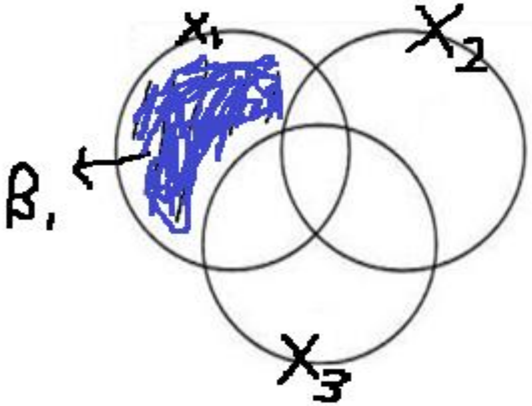
β :?

Ex) $e(x_1|1,x_2,x_3)$ is orthogonal to span$\{1,x_2,x_3\}$.

$$\text{span}\{1,x_1,x_2,x_3\} = \text{span}\{1,x_2,x_3\} \bigoplus \text{span}\{e(x_1|1,x_2,x_3)\}$$

$$\Rightarrow \hat{y}(1,x_1,x_2,x_3) = \hat{y}(1,x_2,x_3) + \hat{y}(e(x_1|1,x_2,x_3)).$$

$$= \hat{y}(1,x_2,x_3) + \hat{\beta}_1 e(x_1|1,x_2,x_3)$$



4 | 5

고유값과 마르코브 체인

시간에 따른 상태의 변화

미래의 상태는 오직 현재에만 의존

Ex)

$$A = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}$$

$$v = c_1 x_1 + c_2 x_2$$

$$A^k v = c_1 (1)^k x_1 + c_2 \left(\tfrac{1}{2}\right)^k x_2$$

As k increases, $A^k v$ approaches $c_1 x_1$ = steady state

5 | 6

Spectral clustering

1. Calculate Laplacian Matrix (L = D – A)
2. Spectral decompose L

$$\begin{pmatrix} q_1 \cdots q_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} q_1^T \\ \vdots \\ q_n^T \end{pmatrix} (\lambda_1 \geq \cdots \geq \lambda_n)$$

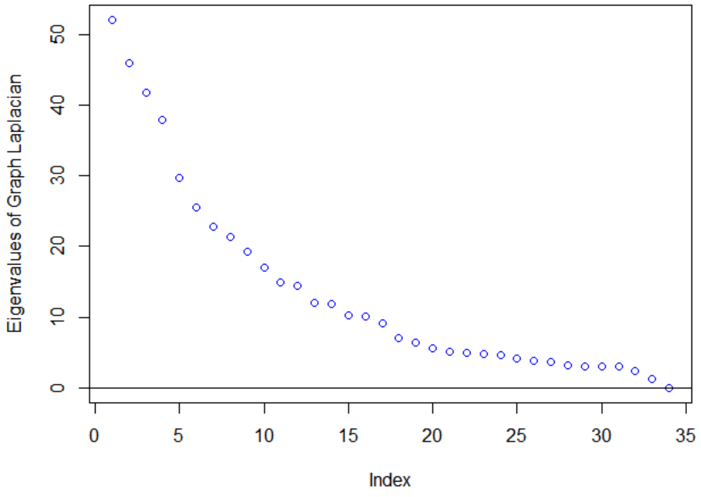
$$\lambda_n = 0 \quad \lambda_{n-m} \approx \cdots \approx \lambda_{n-1} \approx 0$$

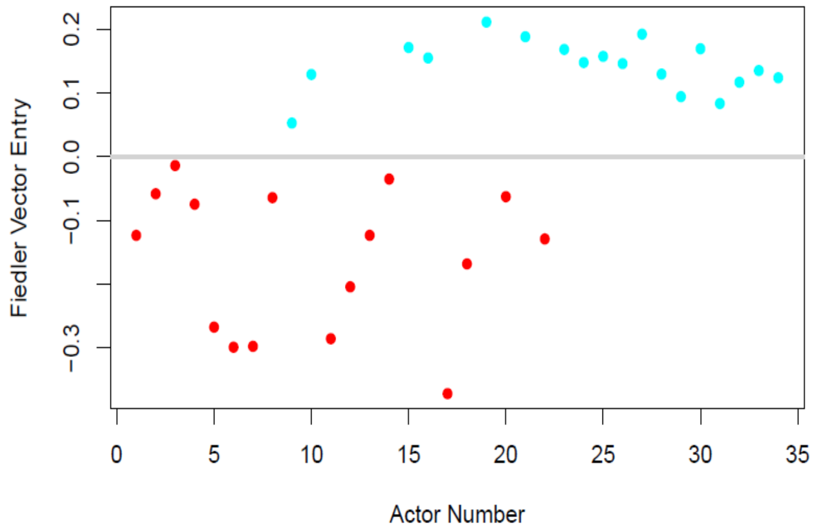
3. Select eigenvectors whose eigenvalue are close to zero

4. Using eigenvectors apply k-means clustering

5 | 6

Spectral clustering





6 | 6

판별분석

$$R(x) = \frac{x^T S x}{x^T x}$$ S : symmetric

Maximum of R(x) : Largest eigenvalue of S

At eigenvector q_1

➡ $$\frac{x^T S x}{x^T M x}$$ S & M : symmetric

Fisher LDA

목표 : 분류를 최대로!!

Maximizes the ratio $$\frac{x^T B x}{x^T W x}$$

B : between group of sum of square

W : within group of sum of square

THANK YOU