Lasso/Ridge Regression
+ Eigenvalue Analysis + PCA

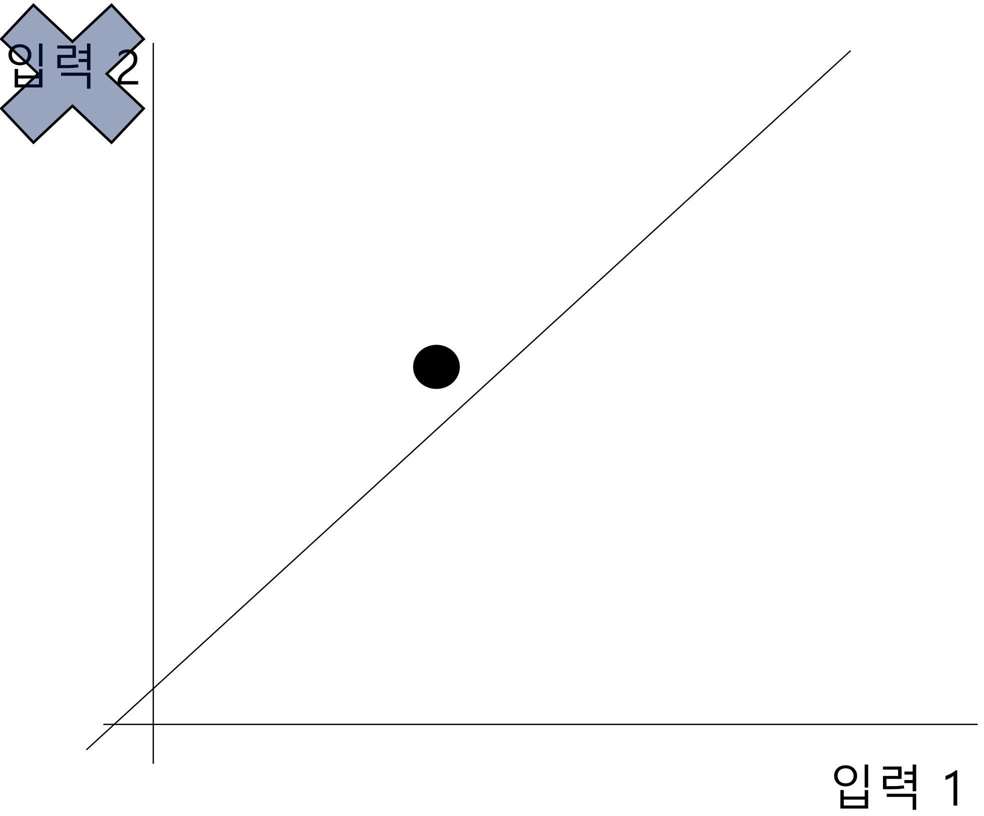
An Introduction to Statistical Learning

황성원

선형모델 확장 이유1: 예측 정확도 향상

선형 모델의 확장!

1. 예측 정확도 향상 2. 모델 해석 용이



$n < p$, 실제 입출력 관계: 선형

분산이 무한대가 된다

~~유일한 최소자승 계수 못 구함~~

Test 관찰 값에서 예측 가능

낮은 바이어스

선형모델 확장 이유2: 모델 해석 용이

선형 모델의 확장!

1. 예측 정확도 향상

2. 모델 해석 용이

표준 선형 모델

$$Y = \beta_0 + \beta_1 X_1 + \cancel{\cdots + \beta_p X_p} + \epsilon$$

입출력 사이의 관계가 단순해져 서로 간의 관계를 더욱 쉽게 파악!

→ 모델 해석 용이!

SW

p(입력 종류 수) 줄이는 방법 (1/3)

1. 부분집합 선택(Subset Selection)

표준 선형 모델

$$Y = \beta_0 + \beta_1 \cancel{X_1} + \cdots + \beta_p X_p + \epsilon$$

출력과 관계 없는 입력 변수X를 없애는 방법으로

최종적으로 줄어든 입력들로 최소자승 Fitting을 수행!

SW

p(입력 종류 수) 줄이는 방법 (2/3)

2. Shrinkage or Regularization

표준 선형 모델

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$$

1. 부분집합 선택
(Subset Selection)

계수를 줄이거나, 0으로 정확히(최소자승법으론 불가능) 수렴 시킨다

모든 p개의 입력들로 Fitting을 수행하지만, 최소자승이 아닌 다른 형태를 사용!

SW

p(입력 종류 수) 줄이는 방법 (3/3)

3. 차원 축소 (Dimension Reduction)

표준 선형 모델

p차원 입력 공간

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$$

M차원 입력 공간

$$Y = \beta_0 + \beta_1 V_1 + \ldots + \beta_M V_M + \varepsilon$$

2. Shrinkage or Regularization

Ridge Regression (L2 Regularization)

1. 일반 선형 회귀 에러

$$\mathrm{RSS} = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$$

2. Ridge 회귀 에러

$$\mathrm{RSS} + \boxed{\lambda \sum_{j=1}^{p} \beta_j^2}$$

Lasso Regression (L1 Regularization)

1. 일반 선형 회귀 에러

$$\mathrm{RSS} = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2$$

2. Lasso 회귀 에러

$$\mathrm{RSS} + \boxed{\lambda \sum_{j=1}^{p}|\beta_j|}$$

Another Formulation for Lasso/Ridge

- 제한조건 -

RSS

1. Lasso $$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \right\}$$ **With** $$\boxed{\sum_{j=1}^{p} |\beta_j| \leq s}$$

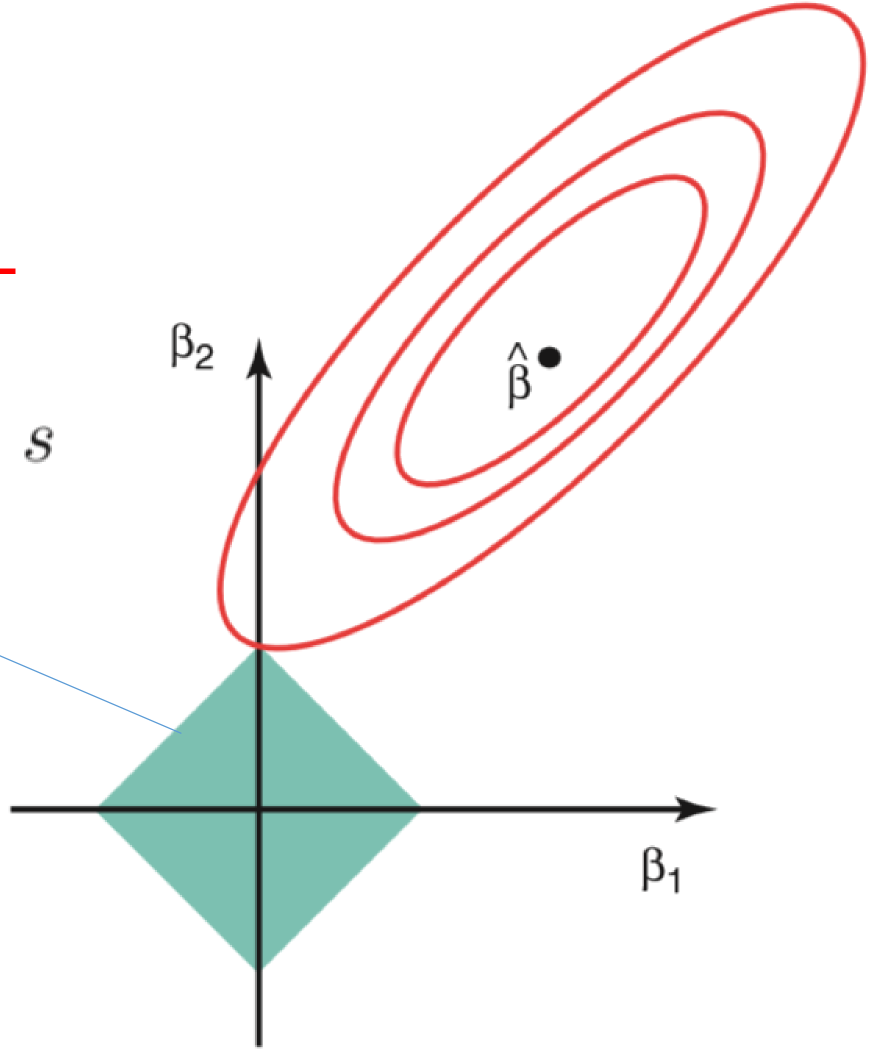
RSS

2. Ridge $$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \right\}$$ **With** $$\boxed{\sum_{j=1}^{p} \beta_j^2 \leq s}$$

Intuition for Lasso/Ridge regression

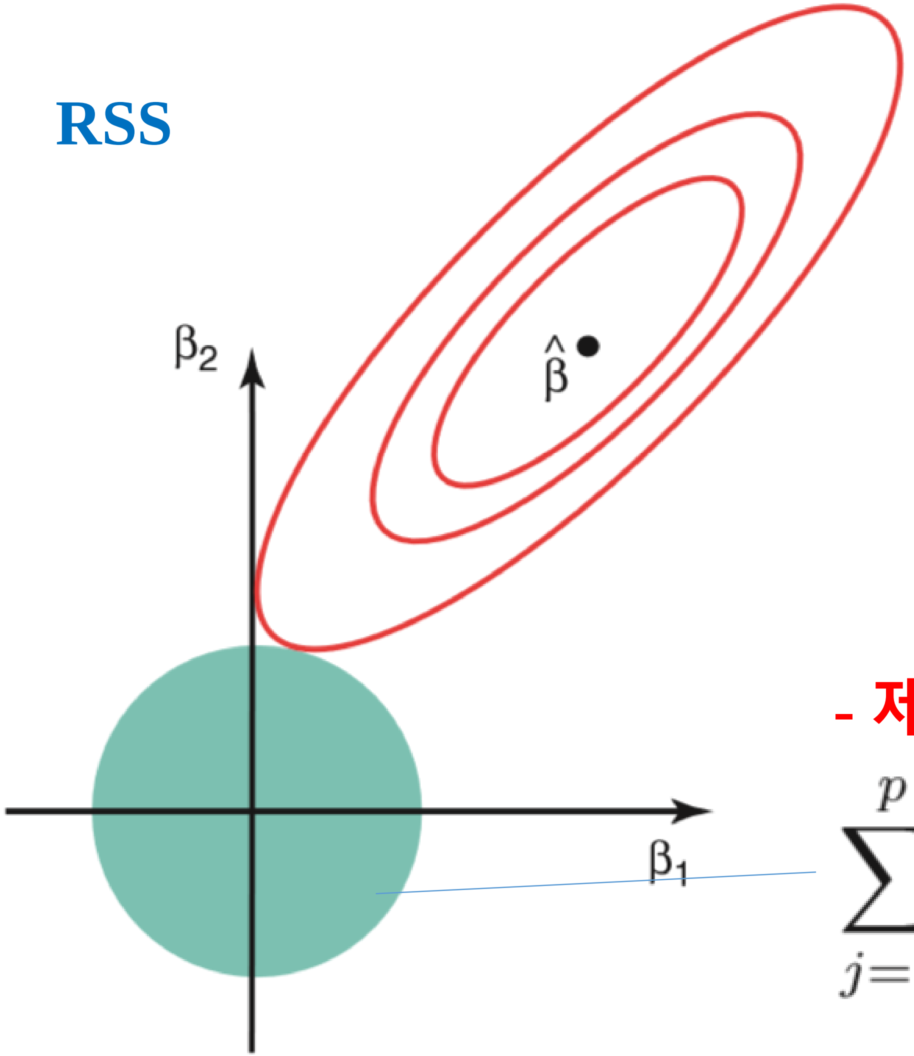
- 제한조건 -

$$\sum_{j=1}^{p}|\beta_j| \leq s$$



1. Lasso

RSS



- 제한조건 -

$$\sum_{j=1}^{p}\beta_j^2 \leq s$$

2. Ridge

Norm, Regularizer L1, L2, …

Norm: 간단하게 벡터의 크기!

$$\left(\sum_i \|\theta_i\|^p\right)^{\frac{1}{p}}$$

$$p = 1$$

$$\sum_{j=1}^{p} |\beta_j| \le s$$

1. Lasso

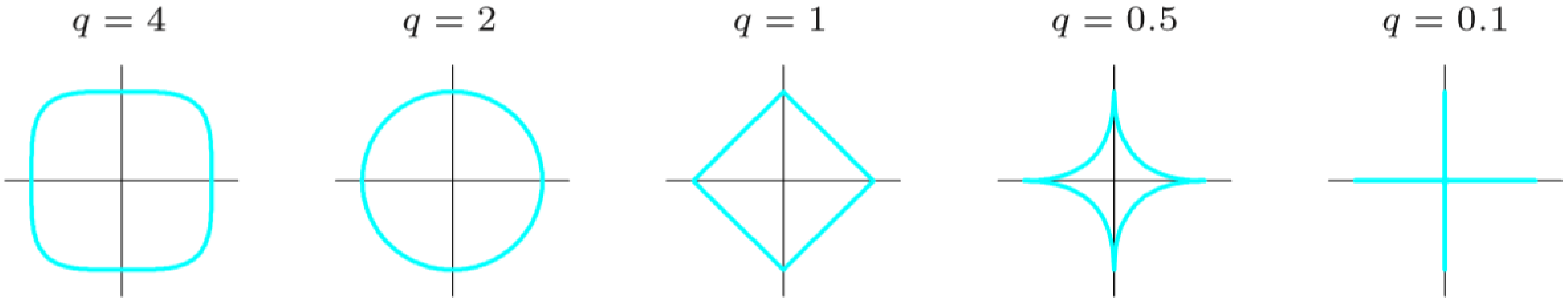
$$p = 2$$

$$\sum_{j=1}^{p} \beta_j^2 \le s$$

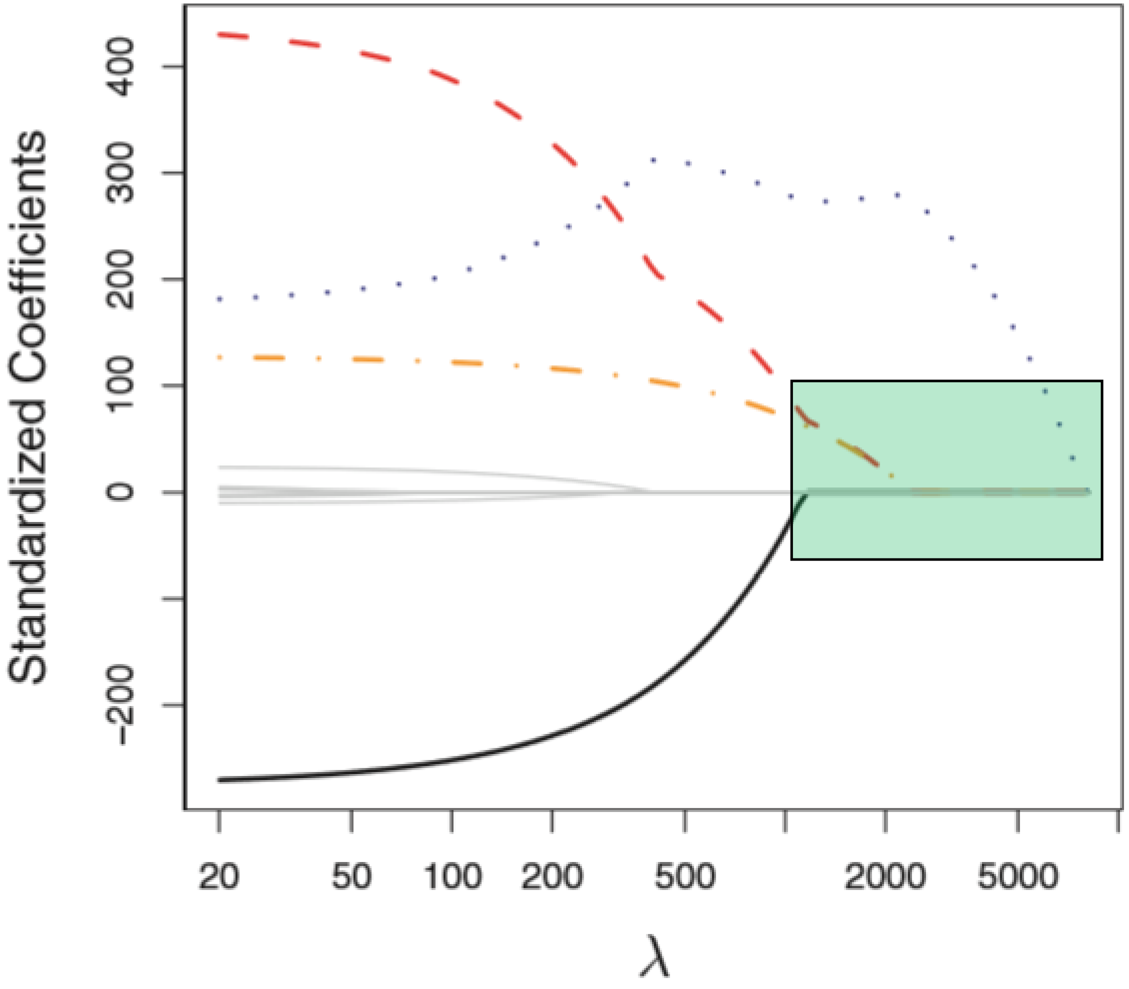
2. Ridge

Extended Version

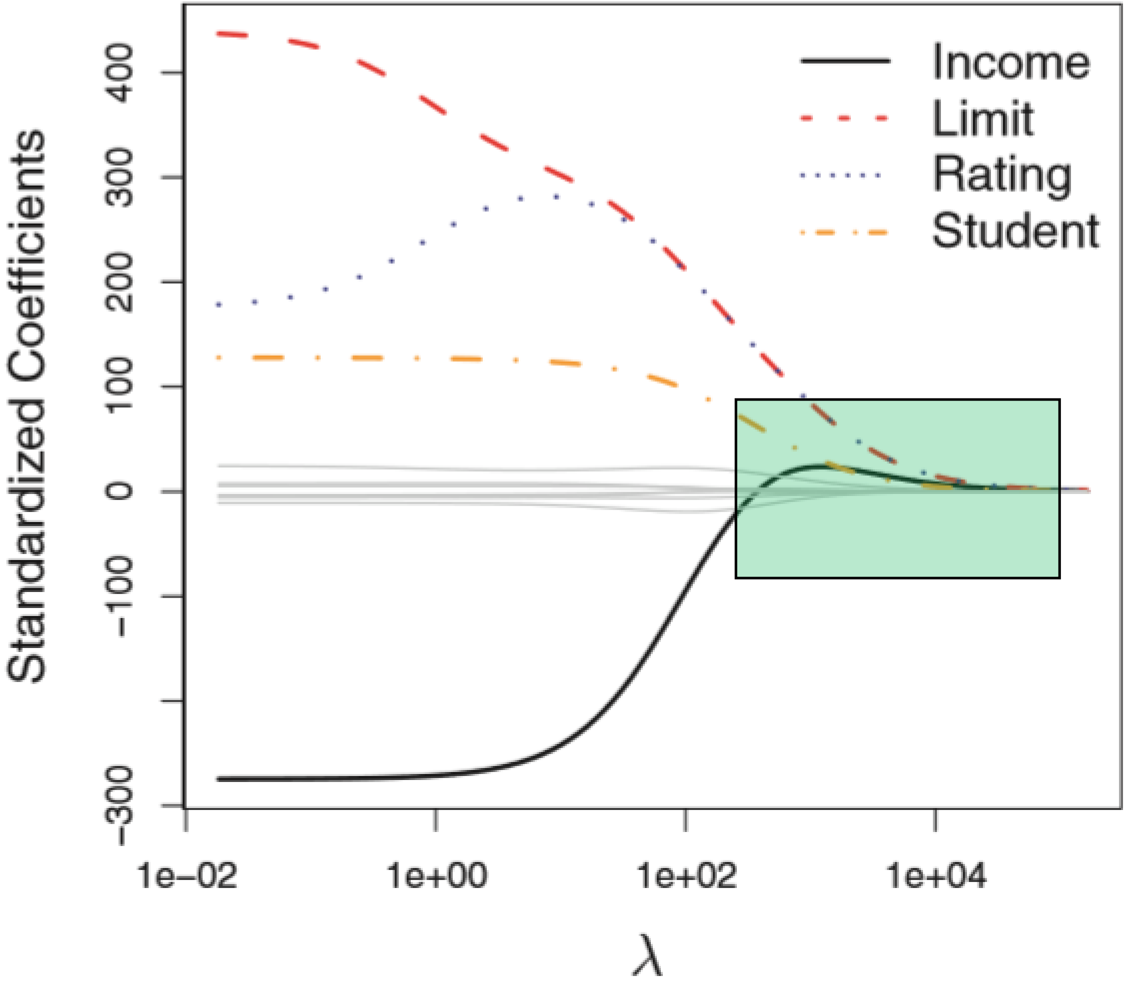
$$\sum_j |\beta_j|^q$$



Ridge VS. Lasso



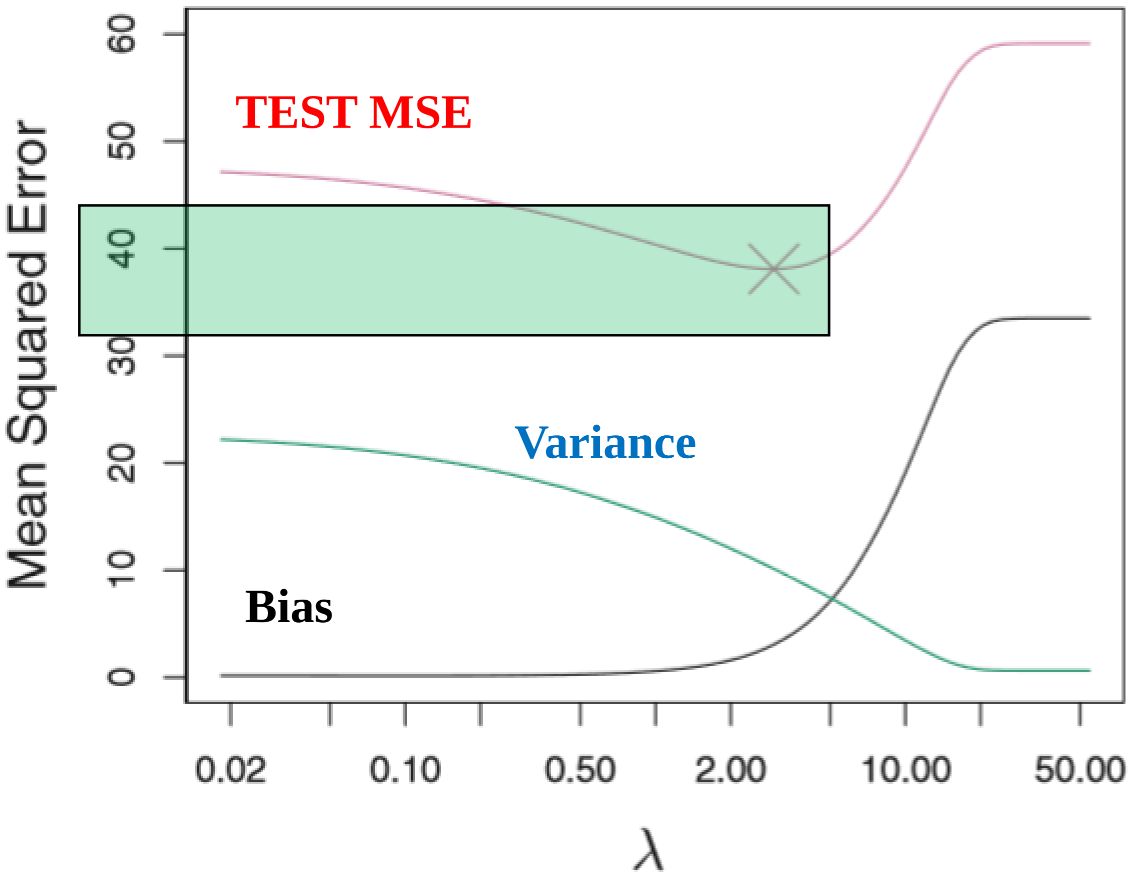
1. Lasso



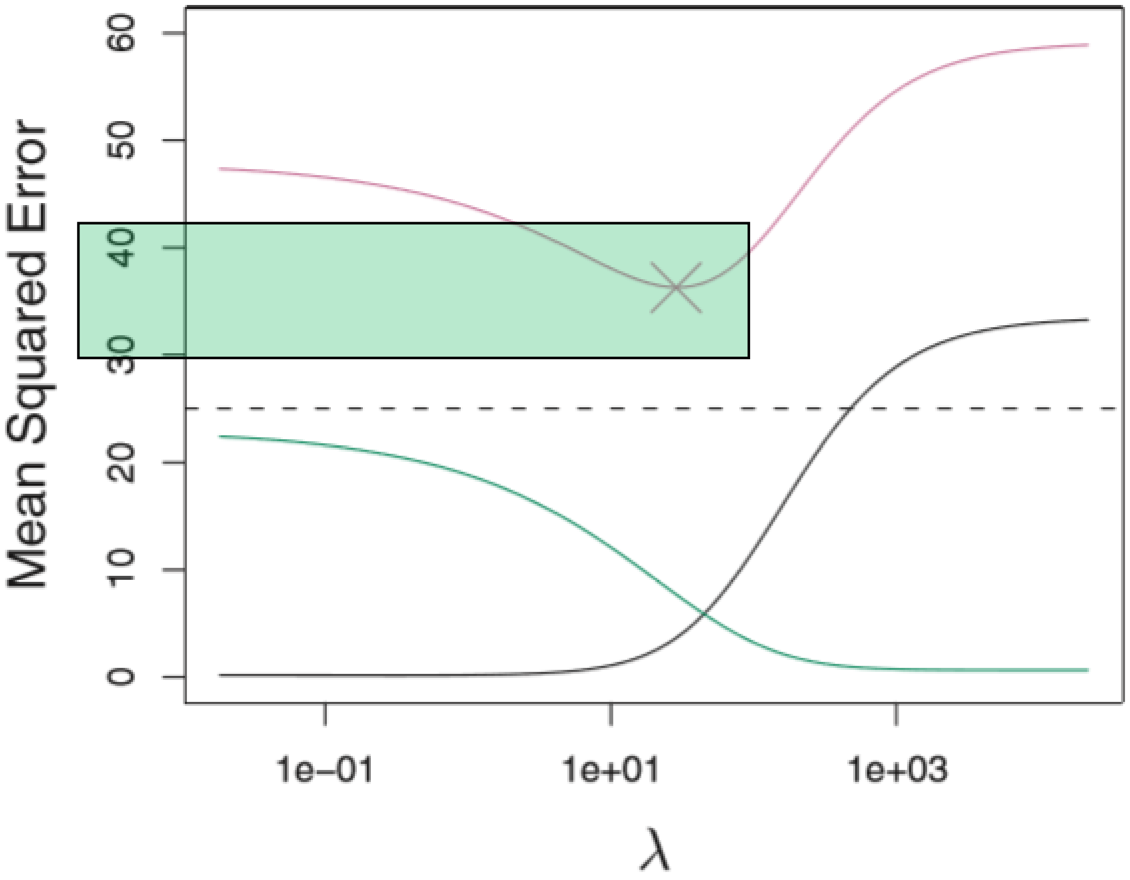
2. Ridge

SW

Lasso VS. Ridge



1. Lasso

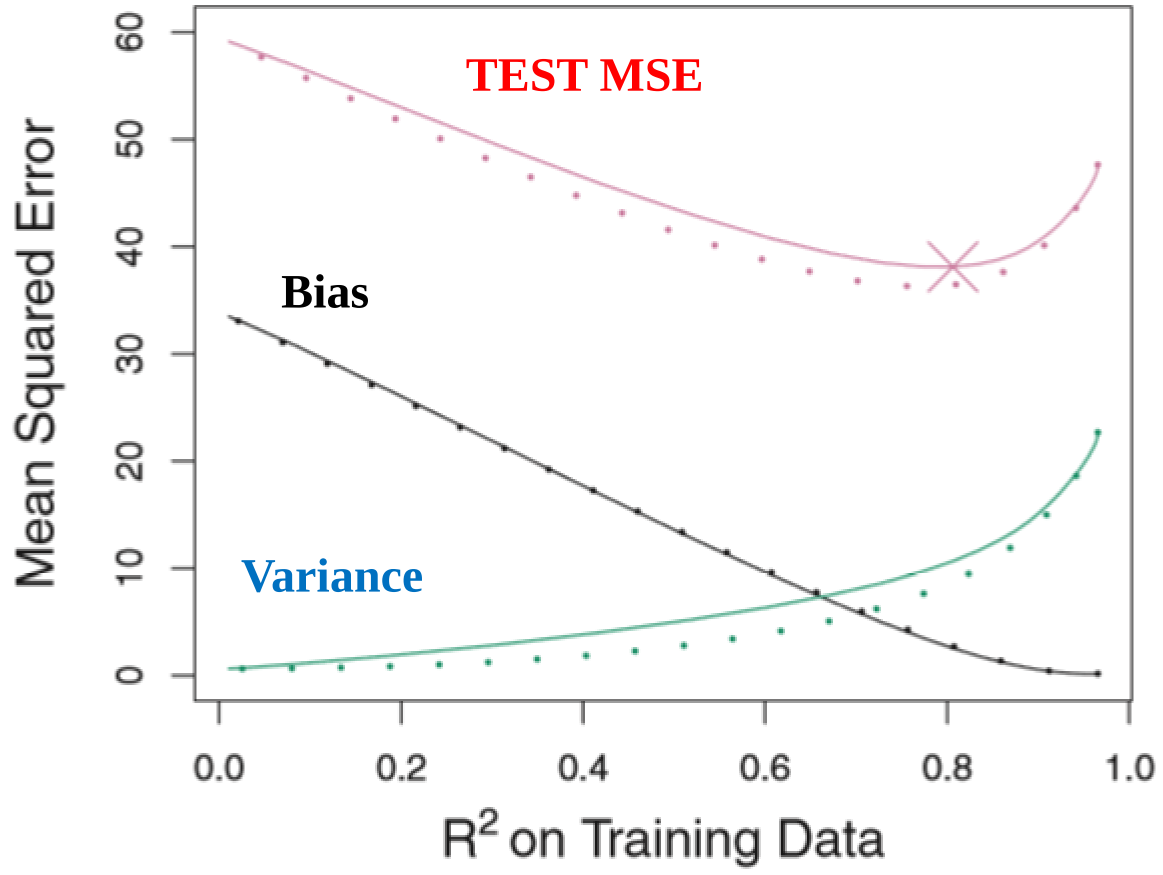


2. Ridge

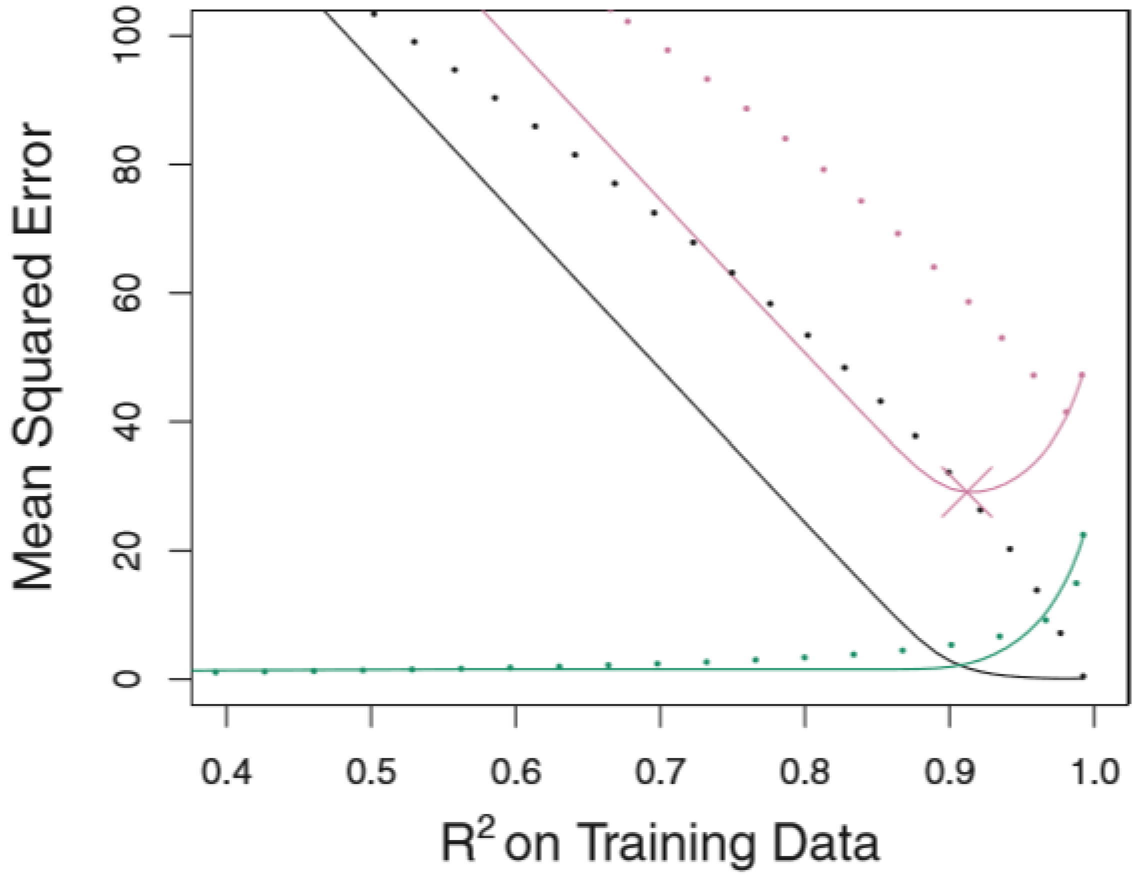
SW

Lasso VS. Ridge

실선 - Lasso
점선 - Ridge



45개 변수가 모두 연관된 경우



45개 변수 중 2개만 연관된 경우

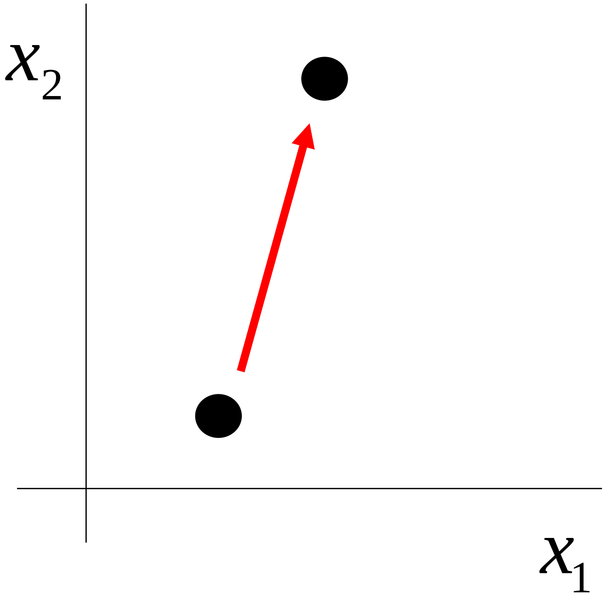
3. Dimension Reduction

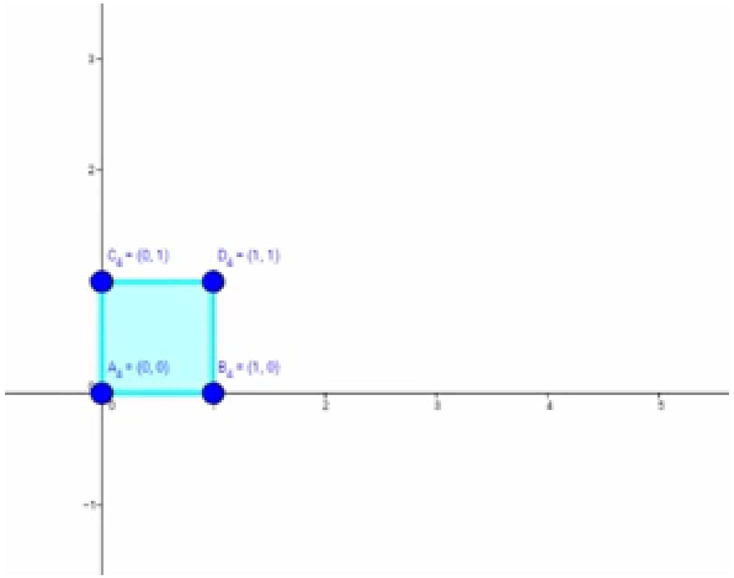
사전 지식 – 고유값, 고유벡터

SW

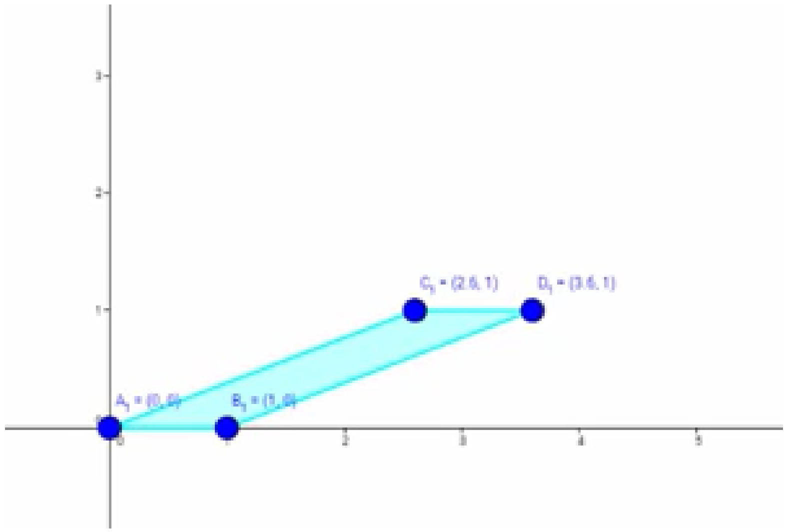
선형 변환 (Linear Transformation)

$$Ax = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x' = \begin{pmatrix} x_1' \\ x_2' \end{pmatrix}$$

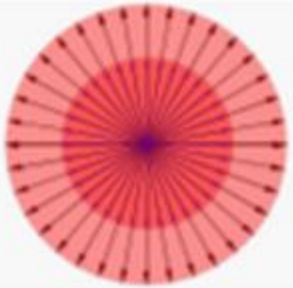
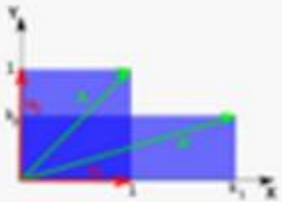
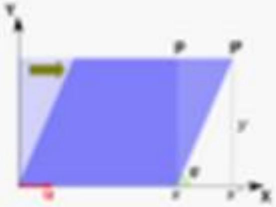




선형 변환



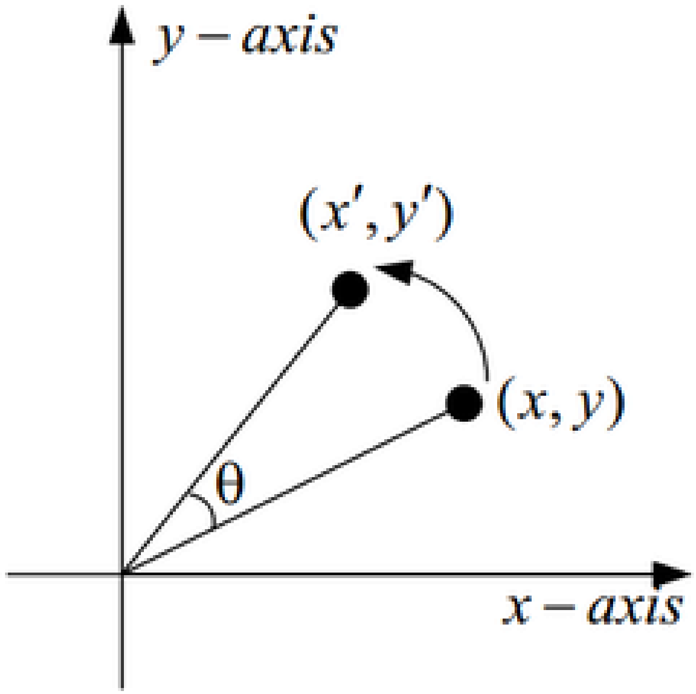
여러 가지 선형 변환

	scaling	unequal scaling	rotation	horizontal shear	hyperbolic rotation
illustration					
matrix	$\begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$	$\begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}$	$\begin{bmatrix} c & -s \\ s & c \end{bmatrix}$ $c = \cos\theta$ $s = \sin\theta$	$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} c & s \\ s & c \end{bmatrix}$ $c = \cosh\varphi$ $s = \sinh\varphi$

$$Ax = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x'$$

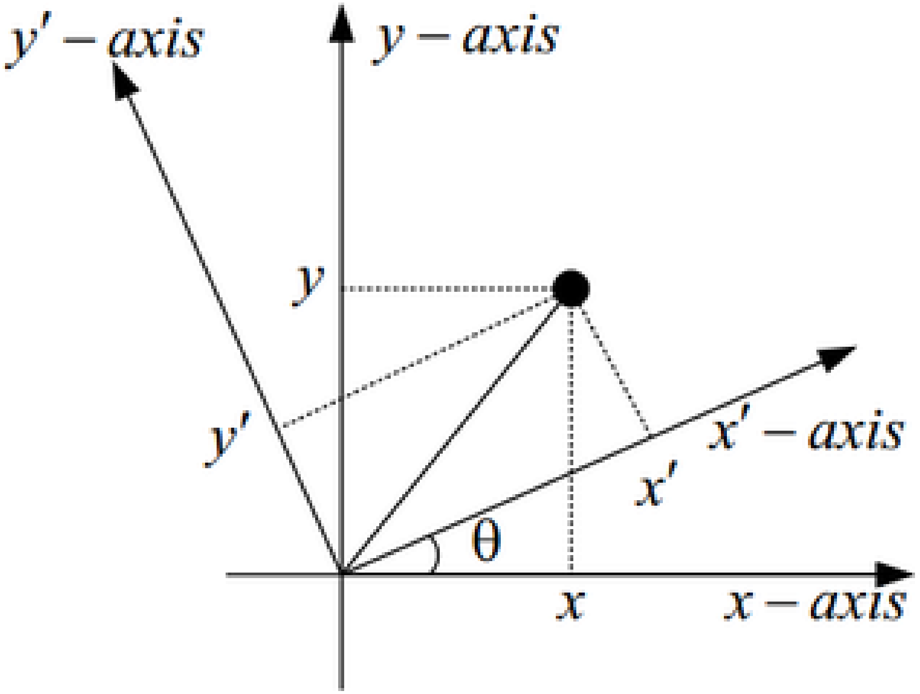
선형 변환 vs. 좌표 변환

■ 점의 회전이동



$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

■ 좌표축의 회전이동

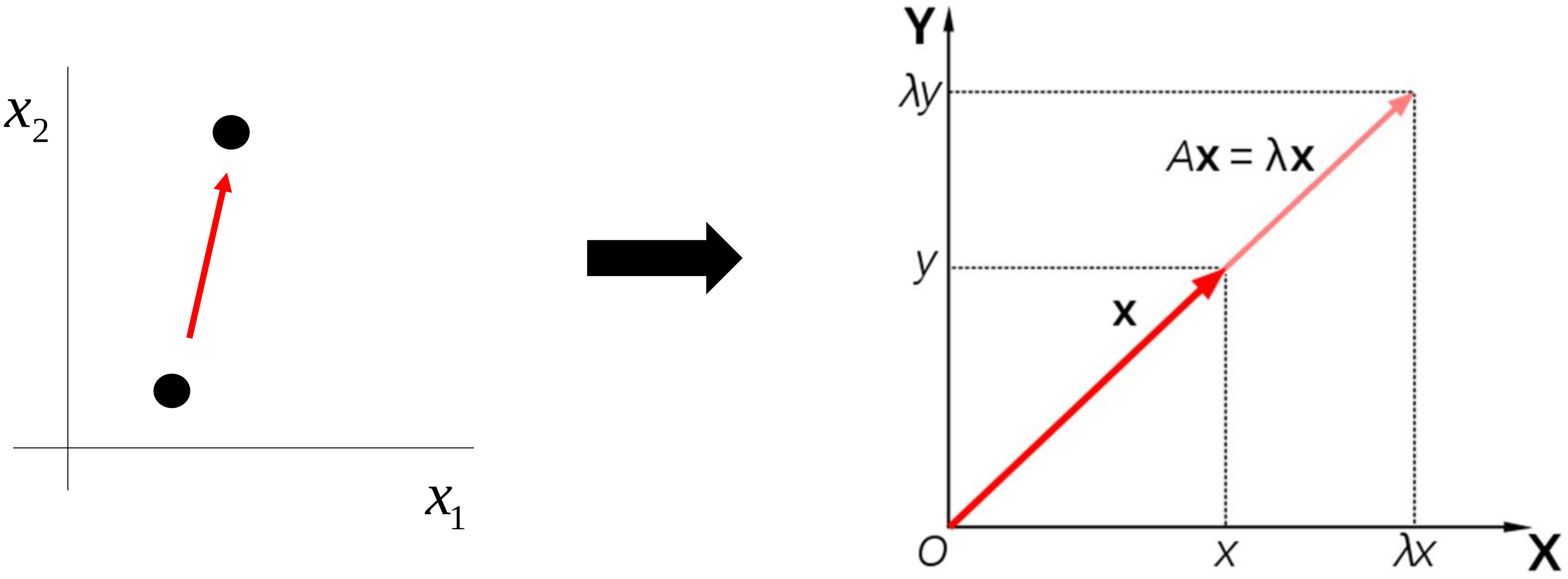


$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

SW

고유 값, 고유 벡터 (Eigenvalue, Eigenvector)

$$Ax = \lambda x$$



Reference for intuition of eigenvector!

2D 기존 원리 바탕으로 한 고유벡터 설명

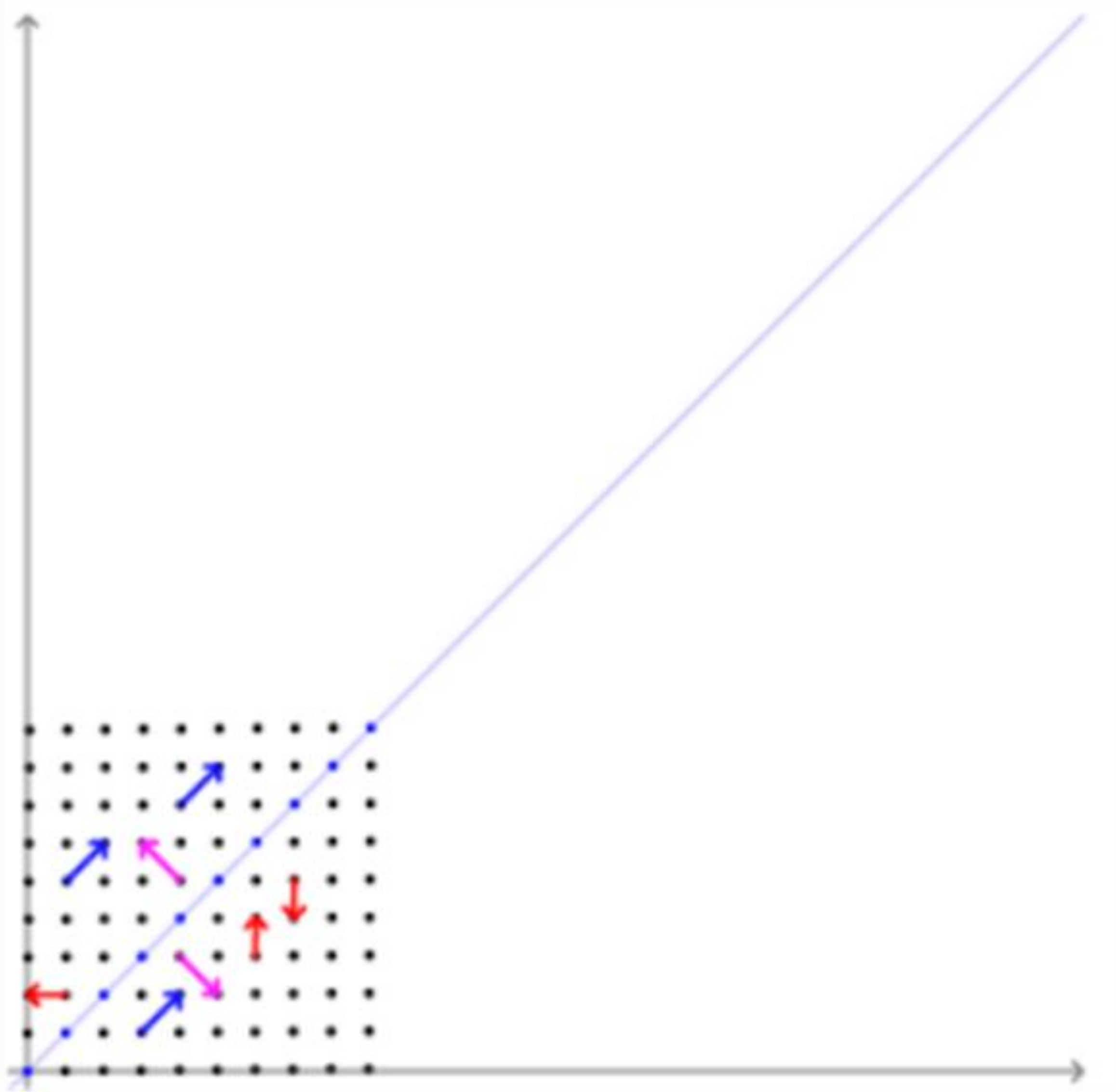
https://www.youtube.com/watch?v=wXCRcnbCsJA

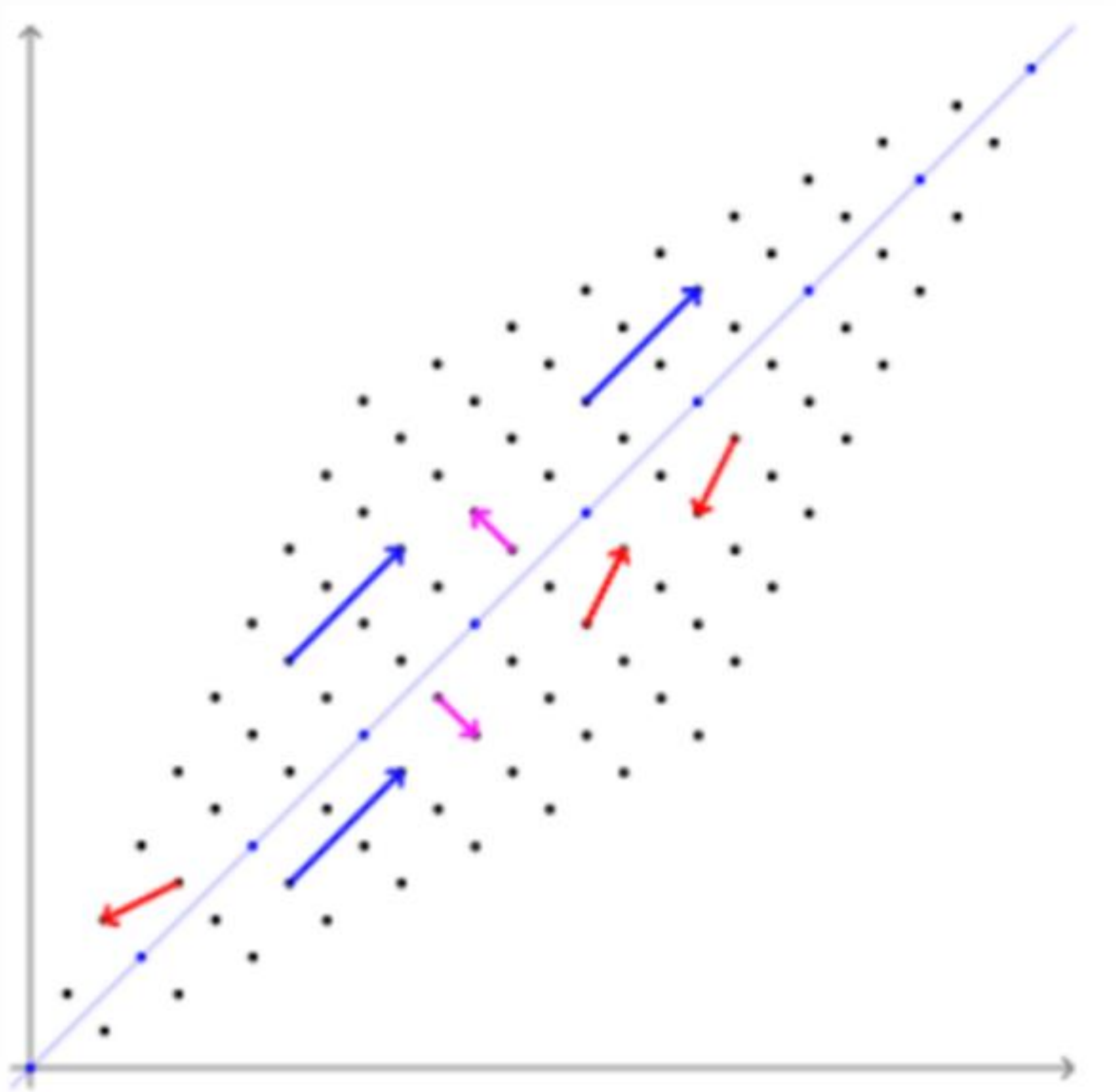
https://www.youtube.com/watch?v=8UX82qVJzYI

3D 예제

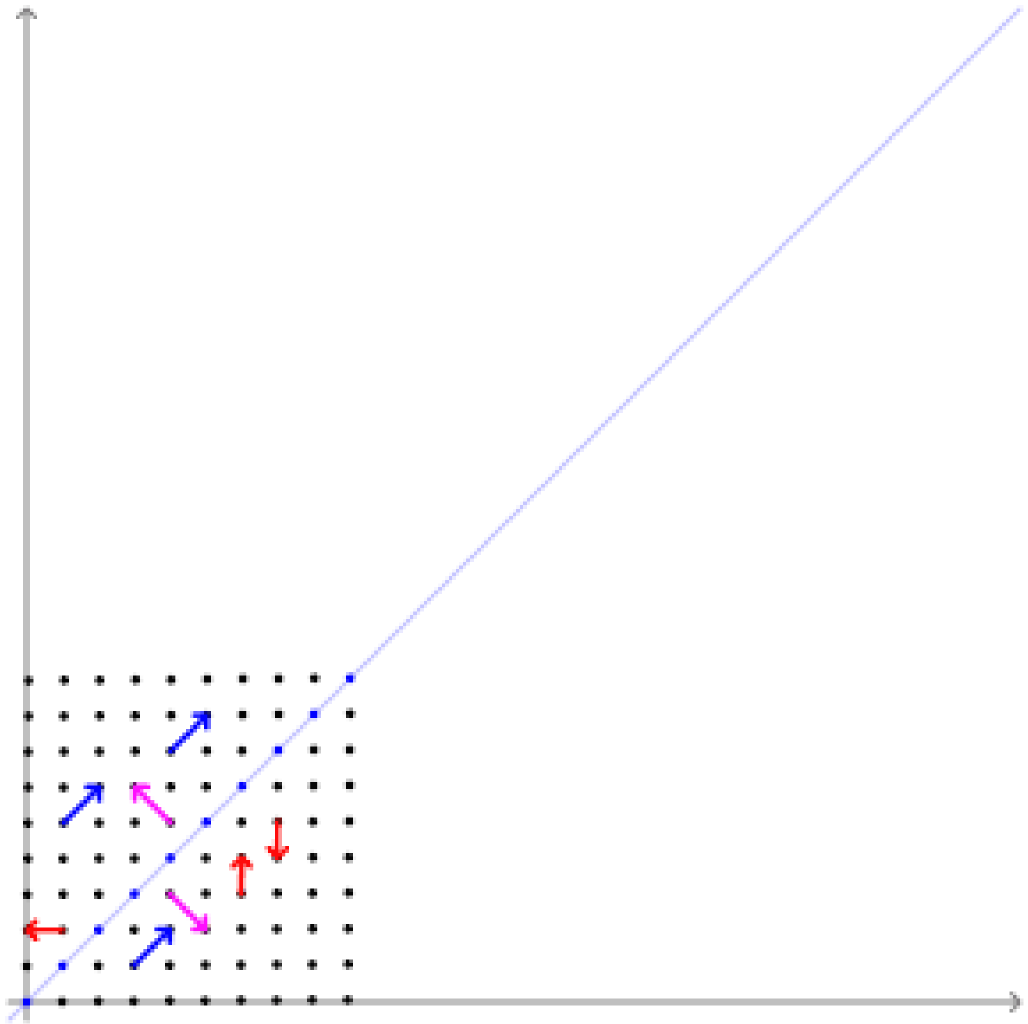
https://www.youtube.com/watch?v=ue3yoeZvt8E

Simple Example – Graphical Explanation





Simple Example – Animation



고유 값, 고유 벡터 (Eigenvalue, Eigenvector)

$$Ax = \lambda x$$

$$(A - \lambda I)x = 0$$

$$\det|A - \lambda I| = 0$$

[역 행렬]

$$\boxed{(A - \lambda I)^{-1} = \frac{1}{\det|A - \lambda I|} \times \ldots}$$

*역 행렬 존재할 경우: Trivial Solution (x=0인 자명해)

따라서, Non-trivial Solution을 찾기 위해 역 행렬 존재 X

SW

Simple Example

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} x = \lambda x$$

$$eigenvalue = \lambda = 1,$$

$$eigenvector = (1, -1)$$

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

$$eigenvalue = \lambda = 3,$$

$$eigenvector = (1, 1)$$

$$(2-\lambda)^2 - 1 = 0$$

$$\lambda = 1, 3$$

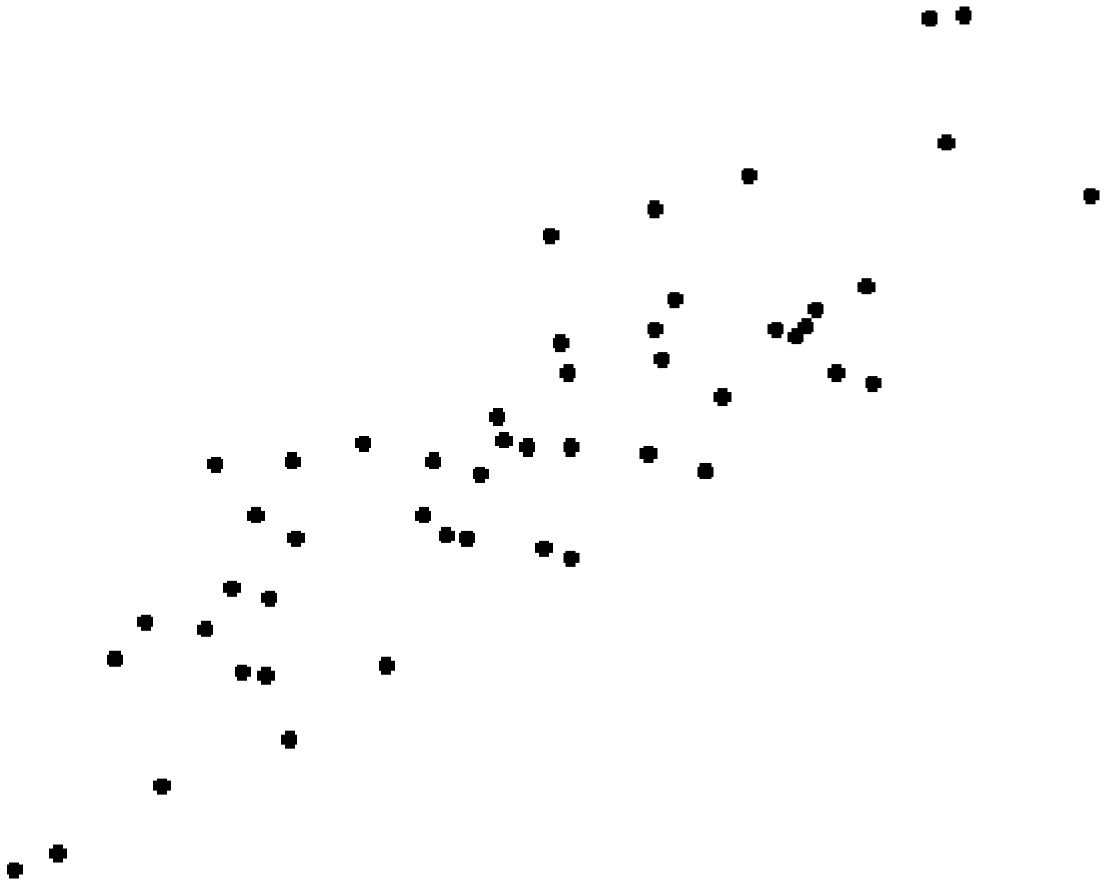
SW

Principle Component Analysis (PCA)

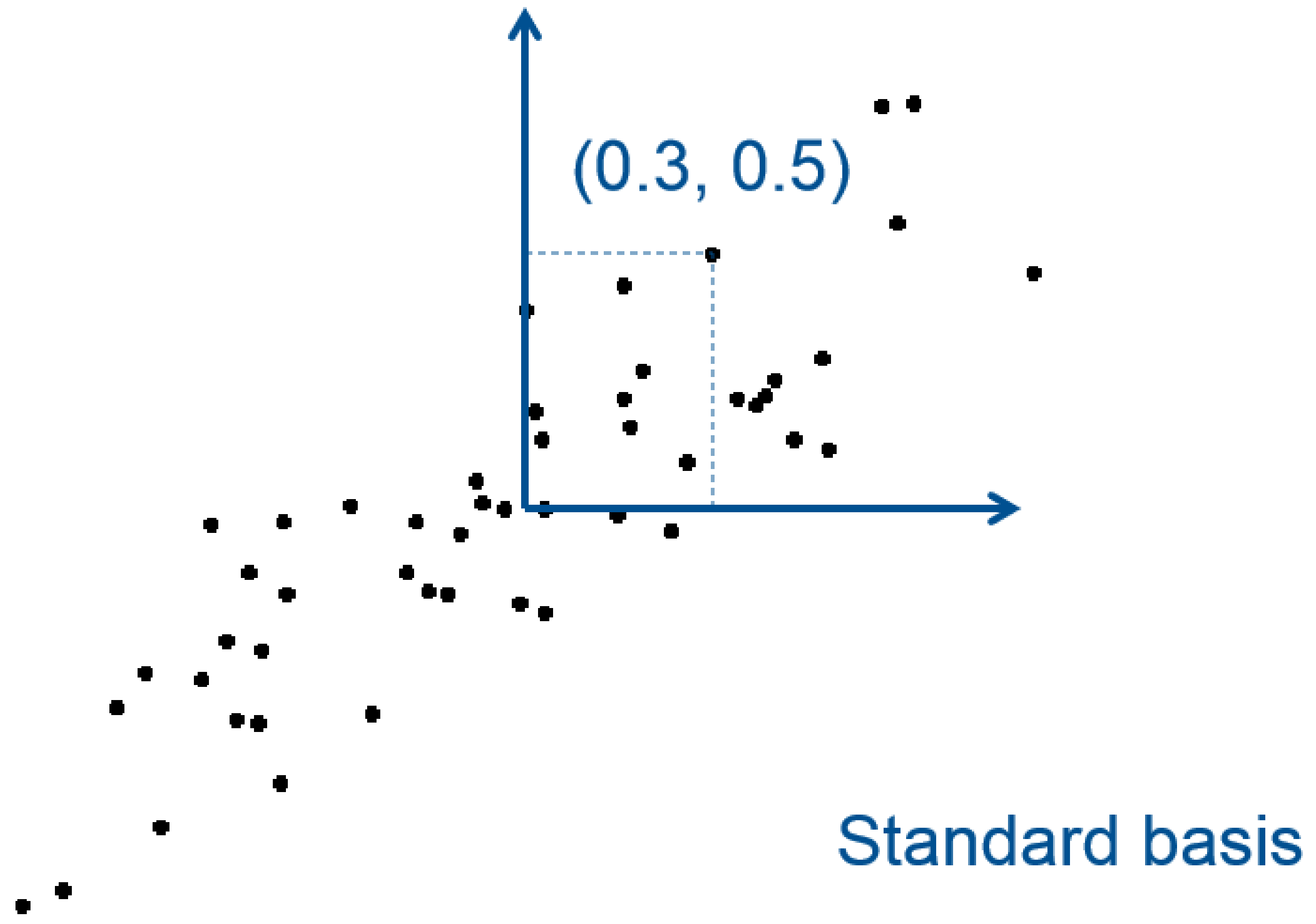
주성분 분석

SW

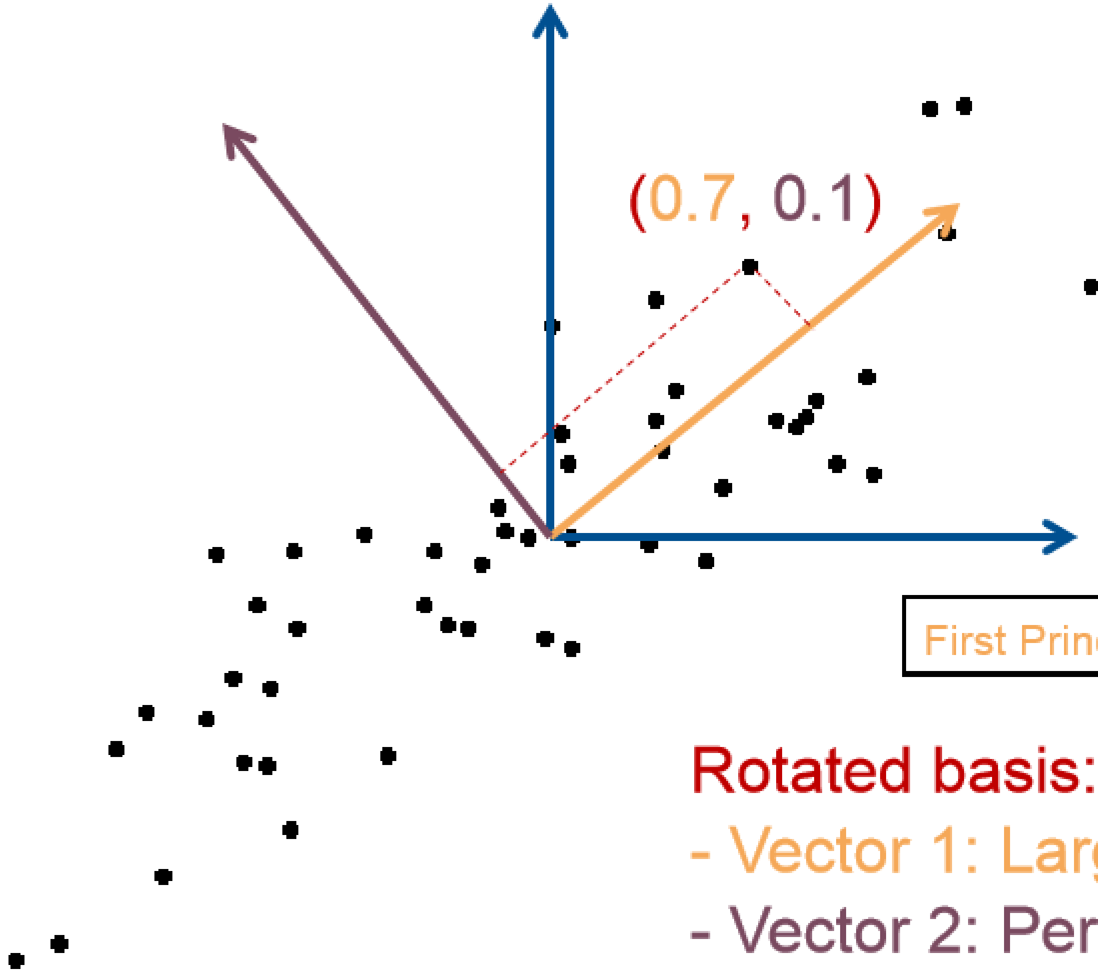
결론



결론



결론



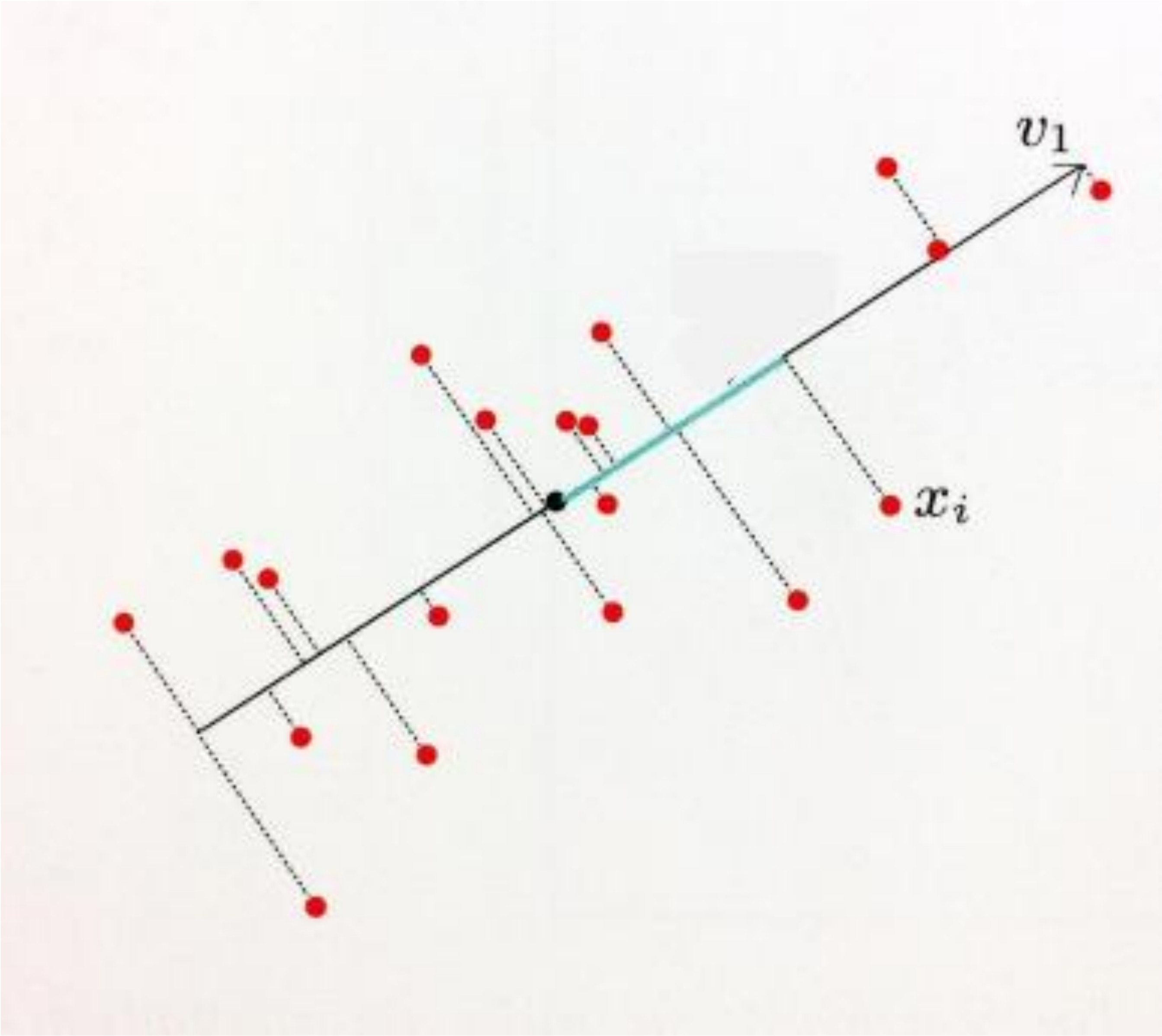
First Principal Component (1.PC)

Rotated basis:

- Vector 1: Largest variance
- Vector 2: Perpendicular

Second Principal Component (2.PC)

결론



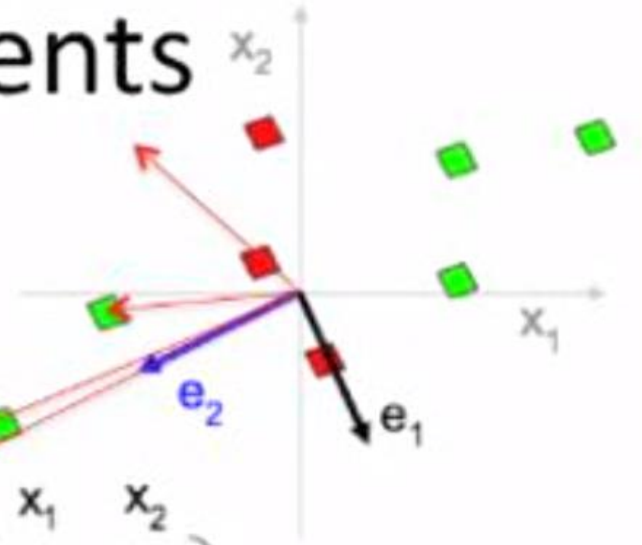
고유벡터에서 행렬A는 PCA에서는?

$$Ax = \lambda x$$

$$\sum x = \begin{pmatrix} \sigma_{x} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{y} \end{pmatrix} x = \lambda x$$

절차

Principal components



- "Center" the data at zero: $x_{i,a} = x_{i,a} - \mu$
  - subtract mean from each attribute
- Compute covariance matrix Σ
  - covariance of dimensions x_1 and x_2:
    - do x_1 and x_2 tend to increase together?
    - or does x_2 decrease as x_1 increases?

$$\begin{array}{c|cc} & x_1 & x_2 \\ \hline x_1 & 2.0 & 0.8 \\ x_2 & 0.8 & 0.6 \end{array}$$

$$\mathrm{var}(a) = \frac{1}{n}\sum_{i=1}^{n} x_{ia}^2$$

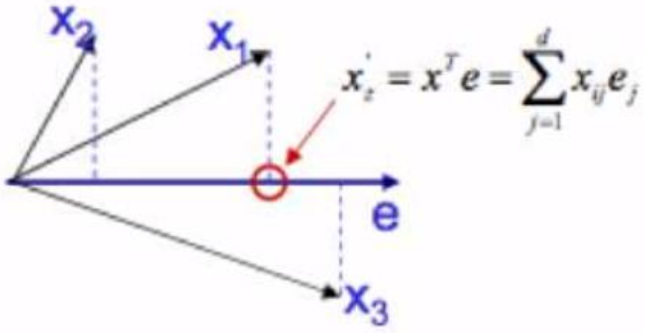
$$\mathrm{cov}(b,a) = \frac{1}{n}\sum_{i=1}^{n} x_{ib} x_{ia}$$

- Multiply a vector by Σ: $\begin{pmatrix} 2.0 & 0.8 \\ 0.8 & 0.6 \end{pmatrix}\begin{pmatrix} -1 \\ +1 \end{pmatrix} \to \begin{pmatrix} -1.2 \\ -0.2 \end{pmatrix}$ again $\to \begin{pmatrix} -2.5 \\ -1.0 \end{pmatrix} \to \begin{pmatrix} -6.0 \\ -2.7 \end{pmatrix} \to \begin{pmatrix} -14.1 \\ -6.4 \end{pmatrix} \to \begin{pmatrix} -33.3 \\ -15.1 \end{pmatrix}$

  slope: 0.400 0.450 0.454 0.454
  - turns towards direction of variance
- Want vectors **e** which aren't turned: $\Sigma\, \mathbf{e} = \lambda\, \mathbf{e}$
  - e … eigenvectors of Σ, λ … corresponding eigenvalues
  - principal components = eigenvectors w. largest eigenvalues

증명

- Select dimension **e** which maximizes the variance
- Points $\boldsymbol{x}_i$ "projected" onto vector **e**:



- Variance of projections: $\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{d} x_{ij}e_j - \mu\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{d} x_{ij}e_j\right)^2$
- Maximize variance
  - want unit length: ||**e**||=1
  - add Lagrange multiplier

$$V = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{d} x_{ij}e_j\right)^2 - \lambda\left(\left(\sum_{k=1}^{d} e_j^2\right) - 1\right)$$

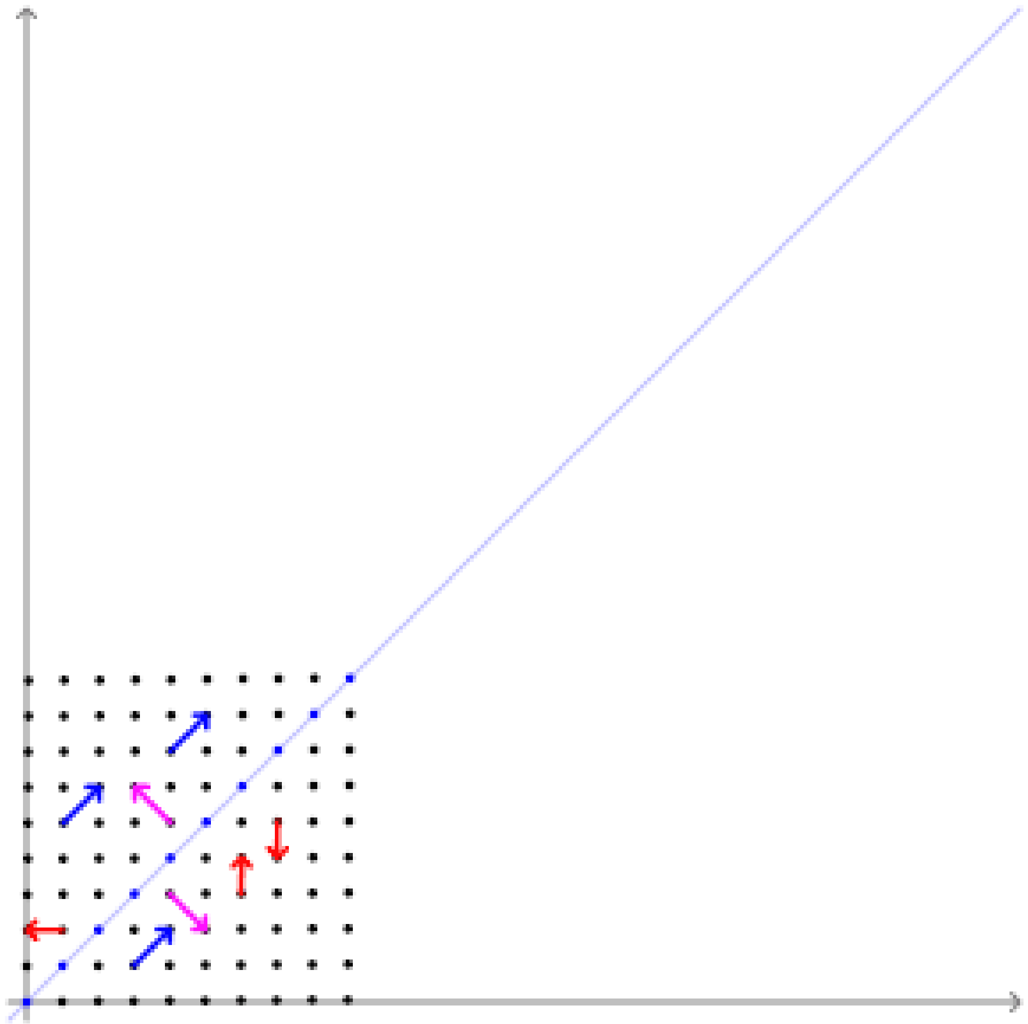
$$\frac{\partial V}{\partial e_a} = \frac{2}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{d} x_{ij}e_j\right)x_{ia} - 2\lambda e_a = 0$$

hold for *a=1..d*

$$2\sum_{j=1}^{d} e_j \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} x_{ia}x_{ij}\right)}_{\text{covariance of a,j}} = 2\lambda e_a$$

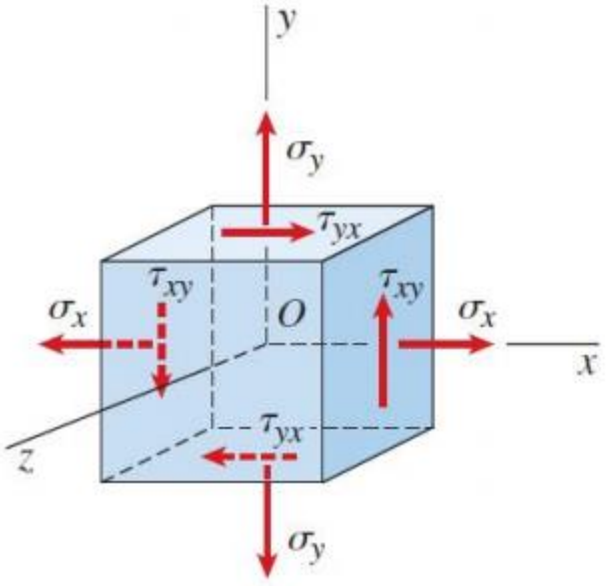
$$\begin{Bmatrix} \sum_{j=1}^{d} \mathrm{cov}(1,j)e_j = \lambda e_1 \\ \vdots \\ \sum_{j=1}^{d} \mathrm{cov}(d,j)e_j = \lambda e_d \end{Bmatrix}$$

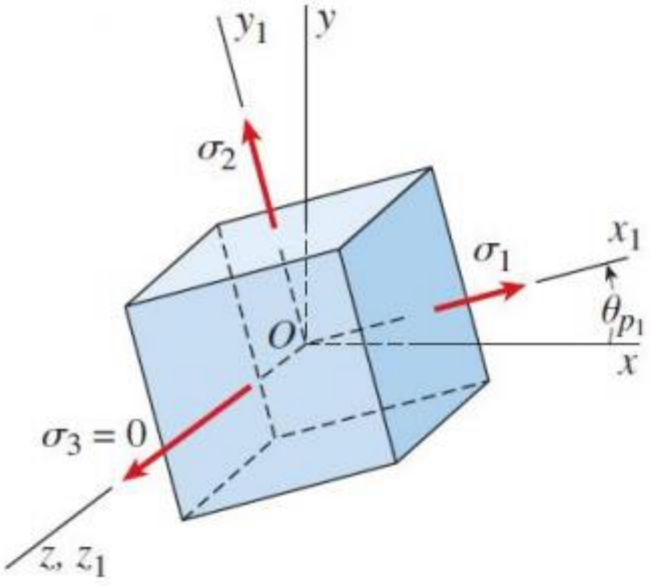
마지막으로 생각할 부분 – 힘과 분산!



마지막으로 생각할 부분 – 주응력

$$A = \begin{pmatrix} \sigma_x & \tau_{xy} & \tau_{xz} \\ \tau_{yx} & \sigma_y & \tau_{yz} \\ \tau_{zx} & \tau_{zy} & \sigma_z \end{pmatrix}$$





SW

Thank you!