

---



## Dimension Reduction의 특성.

1. 차원이 커질수록 노이즈가 줄어 할 확률 증가.

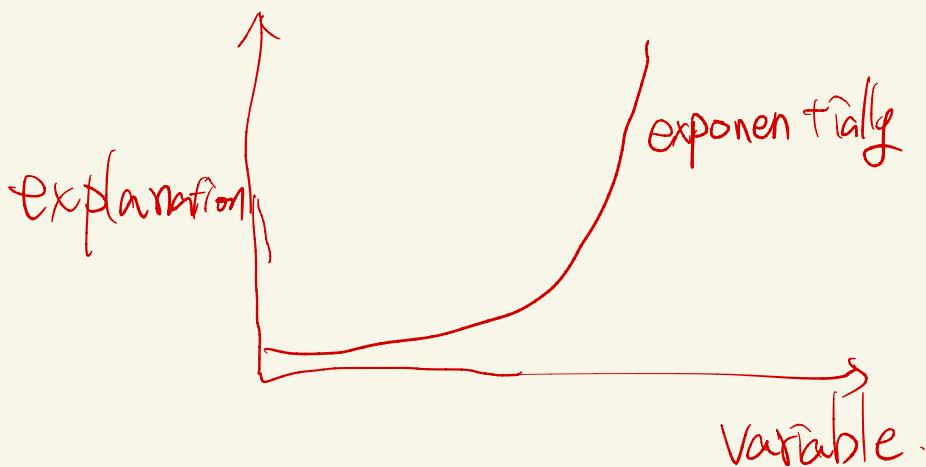
$$y = f(x) + \epsilon$$

$\hookrightarrow$  dimension ↑.  
 $\epsilon \uparrow$

2.  $\Rightarrow$  차원의 증가는 차원에 의하여 진다.

3. 차원의 차수 증가

a) 설명력이遞減, 즉 차원을  
증가할 때 차례로 차감된다.



⇒ Occam's Razor.

$$f(x_1, \dots, x_{10}) \rightsquigarrow R^2_{\text{adj}} = 0.95$$

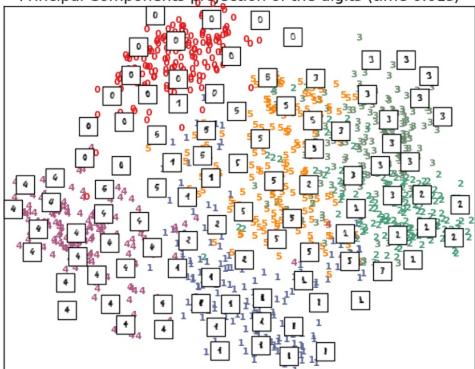
$$f(x_1, \dots, x_3) \rightsquigarrow R^2_{\text{adj}} = 0.95.$$

Win!

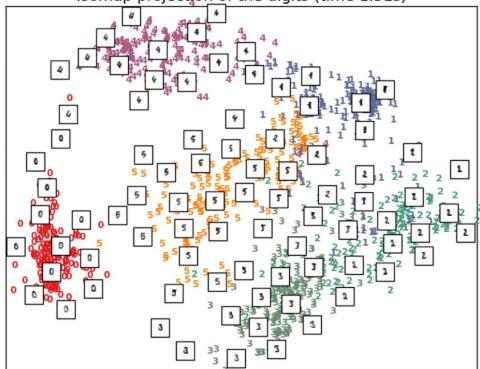
4. 데이터를 표현하는 본질적인 차원 (intrinsic dimension)은 데이터를 표현하는 차원보다 작을 수 있다.

↔

Principal Components projection of the digits (time 0.01s)



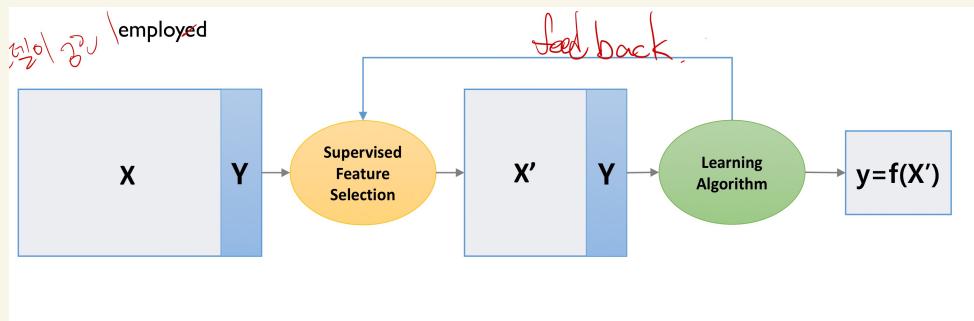
Isomap projection of the digits (time 1.51s)



MNIST Data

⇒ 2차원으로 축소하는 이차원화이다.  
( $16 \times 16 \rightarrow 2$ )

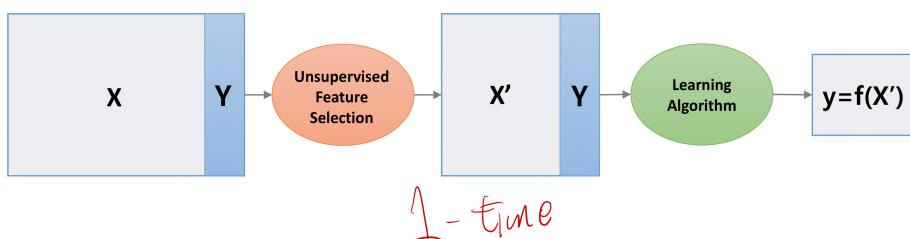
- 일반적으로 훈련 데이터가 추가될 수록 모델의 성능을 증가합니다. (학습률은 어려움)
- Supervised vs Unsupervised Method.
  - Supervised.
    - 훈련 중간에 개선, (feed Back Loop 존재)



## - Unsupervised

· 旣存된 데이터로 X

· 1-time Learning



· feature selection vs extraction

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \xrightarrow{\text{feature selection}} \begin{bmatrix} x_{i_1} \\ x_{i_2} \\ \vdots \\ x_{i_M} \end{bmatrix}$$

↓  
선택된 특징

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \xrightarrow{\text{feature extraction}} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}\right)$$

↓  
제작된 특징

## • Variable Selection Method

### ① Exhaustive Search. (전역 탐색)

- 가능한 모든 변수 조합들을 이용하여 최적의  
변수 집합을 찾는다.

⇒ 즉, 시간이 더욱 오래 걸리는 방법!

=> 사용 X

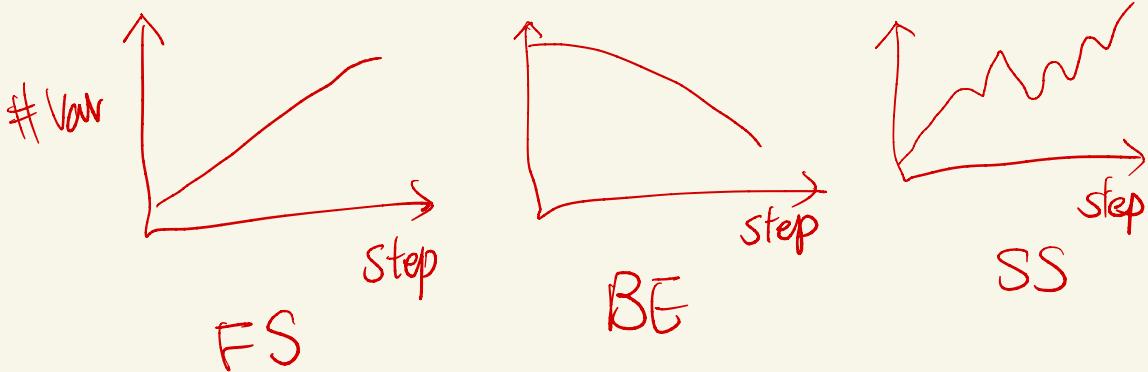
### ② Forward Selection / Backward Elimination

- FS : 변수를 하나씩 증가 시키면서 최적의  
변수 set을 찾는다.

- BE : 변수를 하나씩 줄여면서 최적의 변수 set을  
찾는다.

### ③ Stepwise Selection

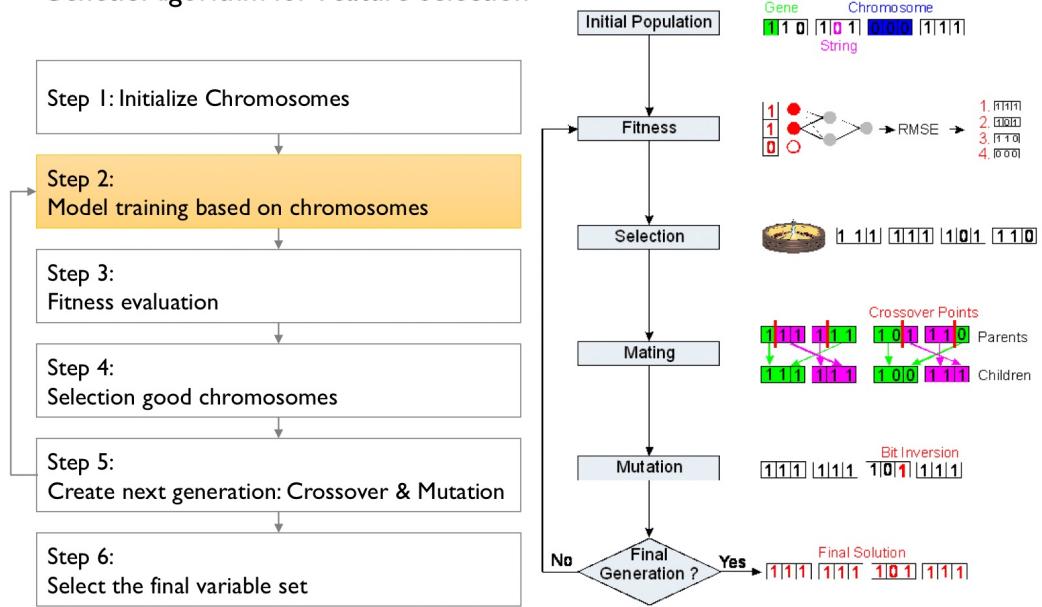
- FS과 BE를 번갈아 가면서 행하여  
최적의 변수 집합을 찾는다.



\* SS 한계는 1, 2 step 까지의 FS 만  
2/3y.

## ④ Genetic Algorithm.

- Genetic Algorithm for Feature Selection



## -Fitness of MLR.

- Adj R<sup>2</sup>
- AIC<sub>c</sub> =  $n \times \ln\left(\frac{SSE}{n}\right) + 2k$
- BIC<sub>c</sub> =  $n \times \ln\left(\frac{SSE}{n}\right) + \frac{2(k+2)n\epsilon^2}{SSE} - \frac{2n^2\epsilon^4}{SSE^2}$

\* n : Data #

k : Vari #

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

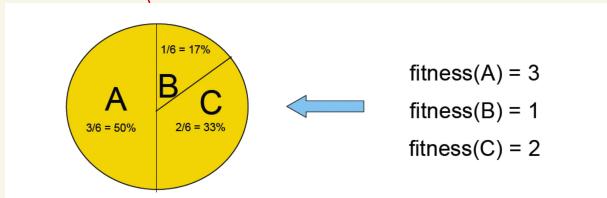
## - Selection

- Determinate Selection

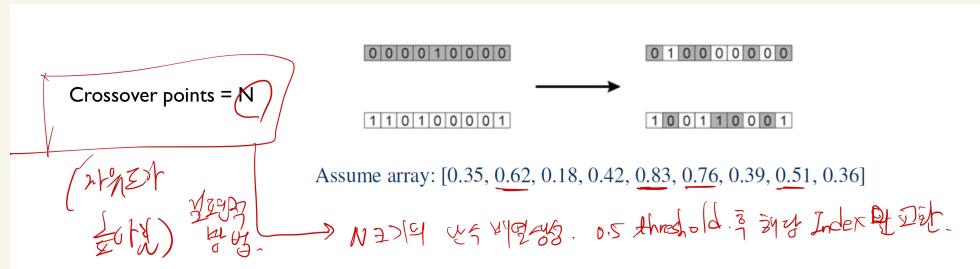
→ top n% of chromosomes

- Probabilistic Selection

⇒ fitness of each selection will be diff.



## - Cross over & Mutation



Mutation ratio : 0.01 (1%).

→ Local Optimum이나くなり기 쉽다!

