Dimesion Reduction의 필요성.

1. 차원이 커질수록 노이즈가 존재 할 확률 증가.

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

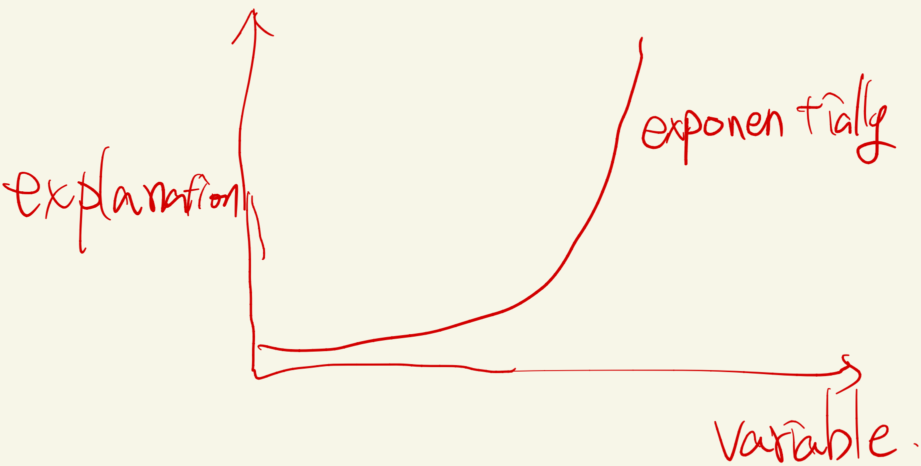
↳ dimension ↑

ε ↑

2. ⇒ 모델의 성능 저하로 이어진다.

3. 차원의 저주 문제

⇒ 설명력이 더 필요, 즉 더 많은 데이터가 필요 하게된다.



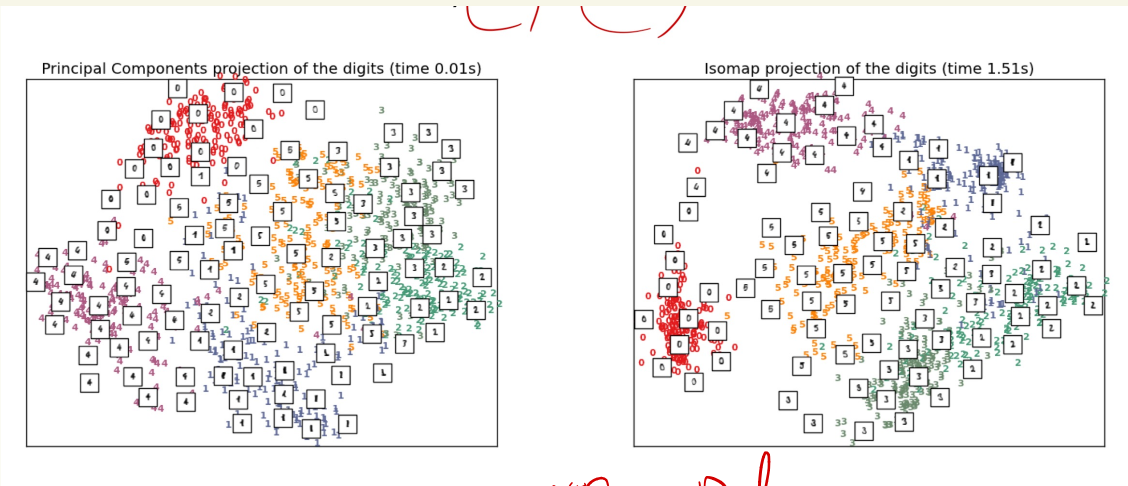
⇒ Occam's Razor.

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

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

Win!!

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



MNIST Data

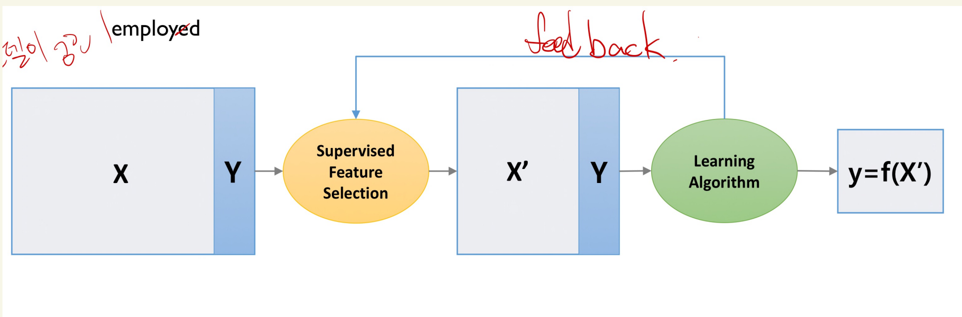
⇒ 2차원으로 축소해도 어느정도 분리가 됨.

(16×16 → 2)

· 일반적으로 독립된 변수가 증가될 수록 모델의 성능은 증가한다. (현실적으로는 어려움.

· Supervised vs Unsupervised Method.

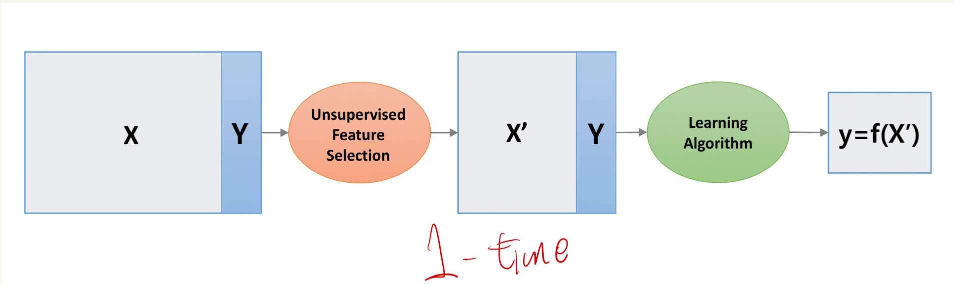
- Supervised.
  · 모델이 중간에 개입. (feed Back Loof 존재)



- Unsupervised

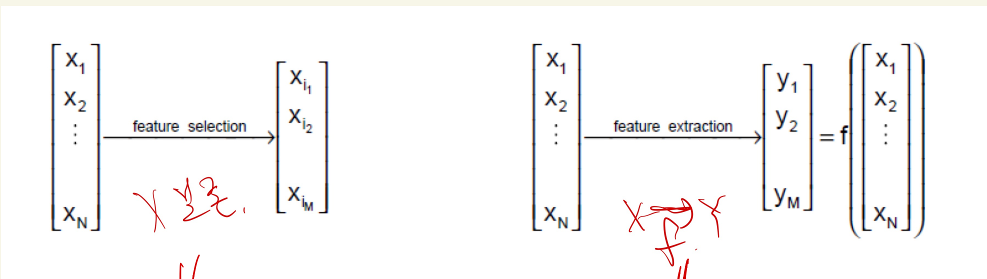
  · 모델이 중간에 개입 X

  · 1-time Learning



1-time

· feature selection vs extraction



X 보존.

$X \xrightarrow{f} Y$

⇓ 부분집합을 선택

⇓ 새로운 부분을 선택.

· Variable Selection Method

① Exhaustive Search (전역탐색)

- 가능한 모든 변수 조합들을 이용하여 최적의 변수 집합을 찾는다.

⇒ 즉, 시간이 너무 오래 걸리는 방법!

⇒ 사용 X

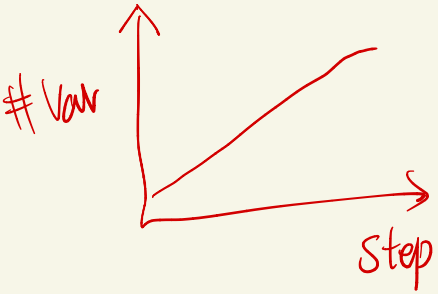
② Forward Selection / Backward Elimination

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

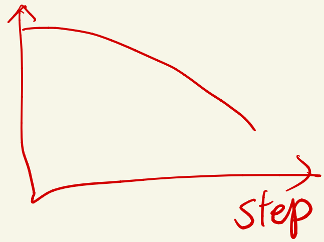
- BE : 변수를 하나씩 줄이면서 최적의 변수 set을 찾는다.

③ Stepwise Selection

- FS 과 BE를 번갈아 가면서 진행하여 최적의 변수 집합을 찾는다.



FS



BE

SS

*SS 만들어줄 시 1,2step 까지는 FS 만 진행.

④ Genetic Algorithm.

- Genetic Algorithm for Feature Selection

Step 1: Initialize Chromosomes

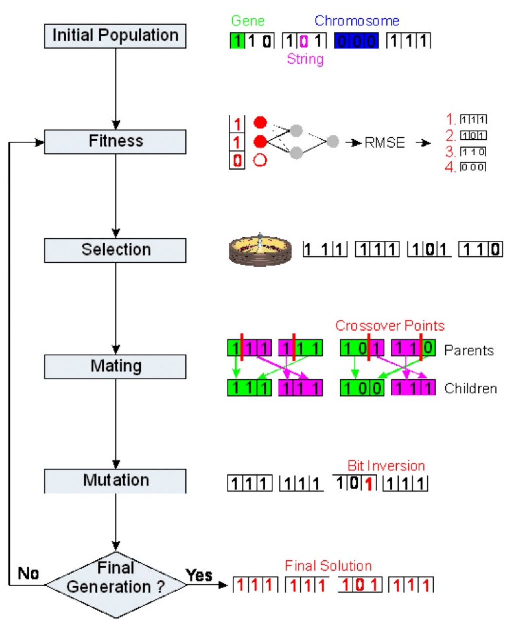
Step 2:
Model training based on chromosomes

Step 3:
Fitness evaluation

Step 4:
Selection good chromosomes

Step 5:
Create next generation: Crossover & Mutation

Step 6:
Select the final variable set



- Fitness of MLR.

  - Adj R^2.
  - AIC $= n \times \ln\left(\frac{SSE}{n}\right) + 2k$
  - BIC $= n \times \ln\left(\frac{SSE}{n}\right) + \frac{2(k+2)n\sigma^2}{SSE} - \frac{2n^2\sigma^4}{SSE^2}$

* n : Data #

k : Vari #

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

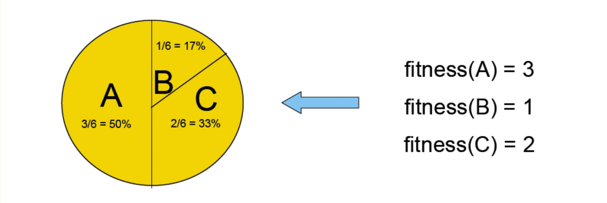
- Selection

  - Determinate Selection

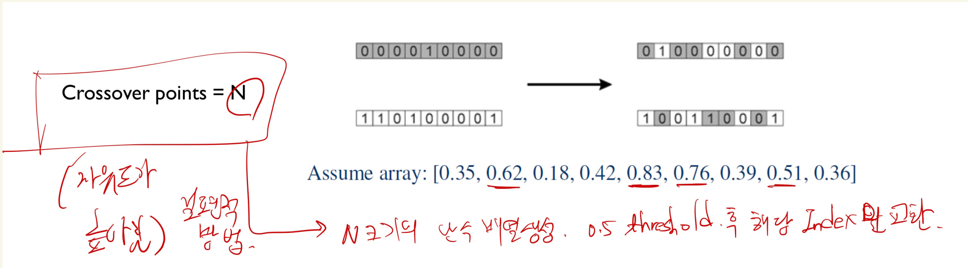
    → top $n\%$ of chromosomes

  - Probabilistic selection

    ⇒ fitness에 따라 selection 확률 부여.



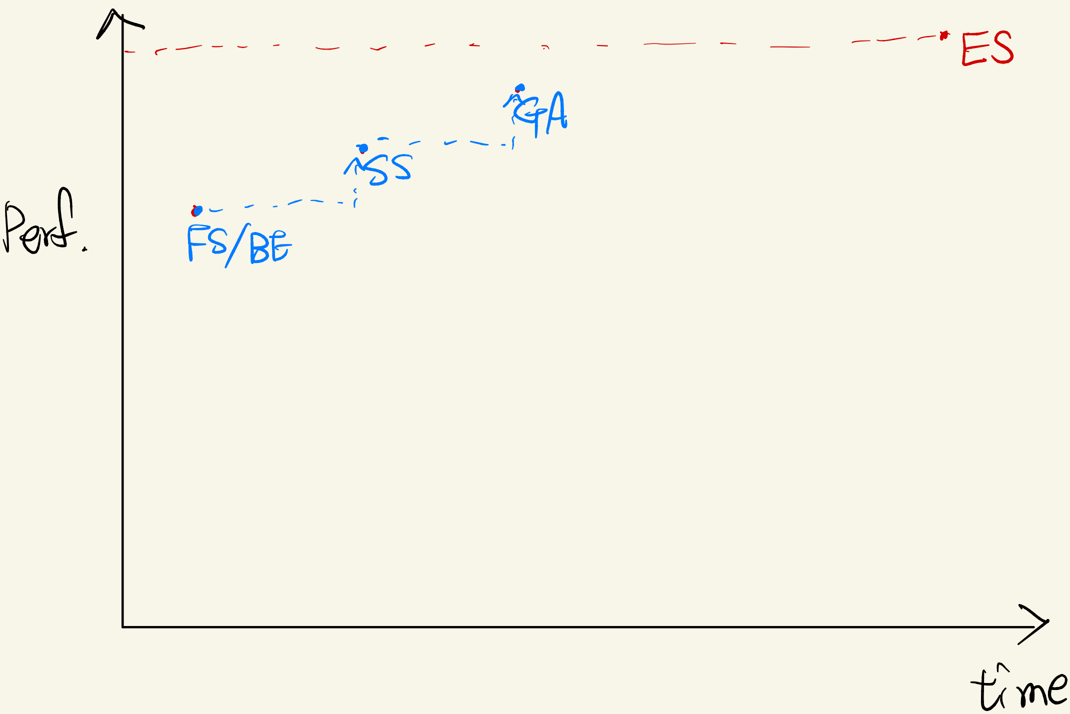
- Cross over & Mutation



(자유도가 높아짐) 보편적 방법. → N 크기의 난수 배열생성. 0.5 threshold 후 해당 Index 값 교환.

Mutation ratio : 0.01 (보통).

⇒ Local Optimum에서 벗어나기 위함!



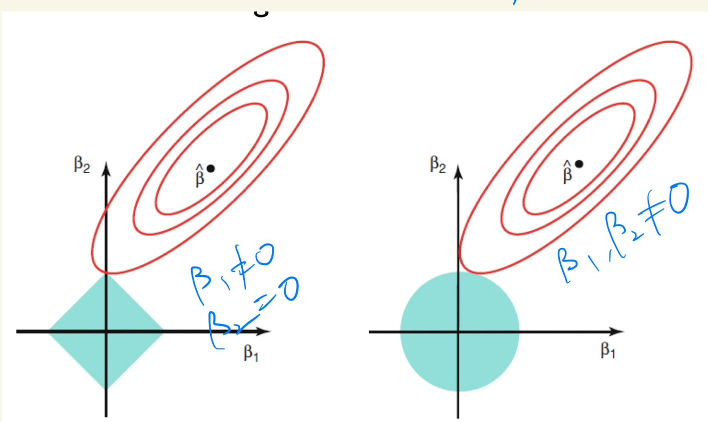
· Shrinkage Method.

① Ridge Regression

- L_2 norm penalty add.
- 각 변수의 coefficients를 줄일 수 있음.
  ⇒ but 0 X, not selection
- correlation ↑ 효과 ↑

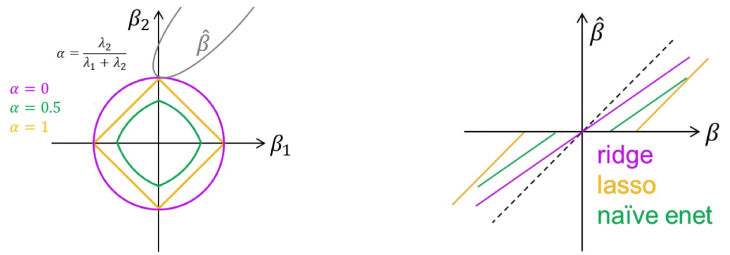
② LASSO.

- L_1 norm penalty add.
- 변수의 selection //



③ Elastic Net.

- Ridge + LASSO.



Ridge	$\hat{\beta} = \min_{\beta} \lvert Y - X\beta \rvert^2 + \lambda_1 \lvert \beta \rvert^2$	shrinkage
Lasso	$\hat{\beta} = \min_{\beta} \lvert Y - X\beta \rvert^2 + \lambda_2 \lvert \beta \rvert^1$	shrinkage, variable selection
Elastic net	$\hat{\beta} = \min_{\beta} \lvert Y - X\beta \rvert^2 + \lambda_2 \lvert \beta \rvert^1 + \lambda_1 \lvert \beta \rvert^2$	shrinkage, variable selection, grouping effect

$\alpha = 0.5 \Rightarrow \lambda_1 = \lambda_2$ 일때.

$\alpha = 0 \Rightarrow \lambda_2 = 0,\ \lambda_1 \neq 0$

$\alpha = 1 \Rightarrow \lambda_1 = 0,\ \lambda_2 \neq 0.$