



Dimension Reduction 의 필요성.

1. 차원이 커질수록 노이즈가 존재 할 확률이 증가.

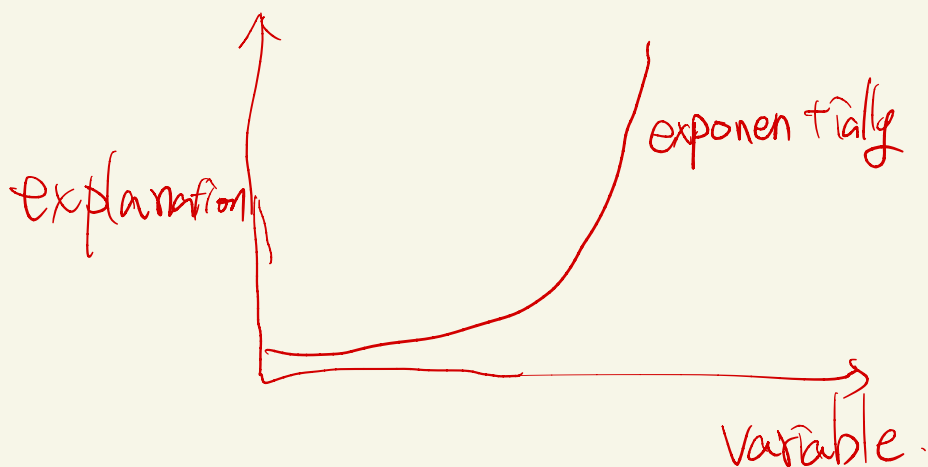
$$y = f(x) + \frac{\epsilon}{L}$$

↳ dimension ↑
ε ↑

2. ⇒ 모델의 성능 저하로 이어진다.

3. 차원의 저주 문제

⇒ 설명력이 저하, 즉 더 많은
데이터가 필요해진다.



⇒ Occam's Razor.

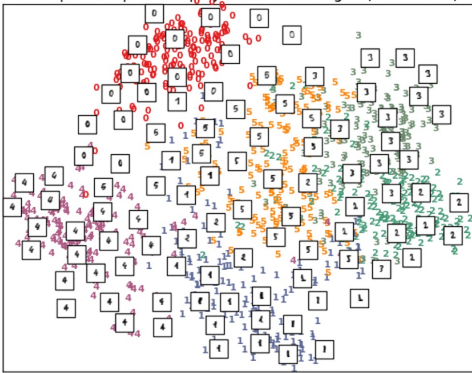
$$f(x_1, \dots, x_{10}) \rightsquigarrow R_{adj}^2 = 0.95$$

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

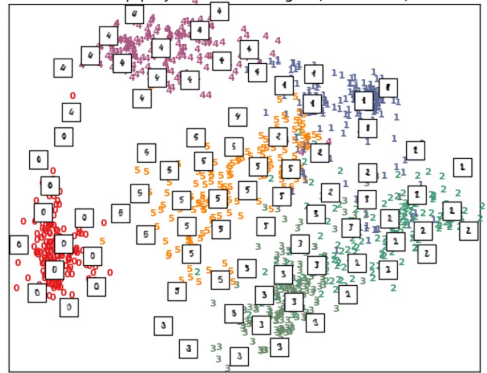
win!

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

Principal Components projection of the digits (time 0.01s)



Isomap projection of the digits (time 1.51s)



MNIST Data

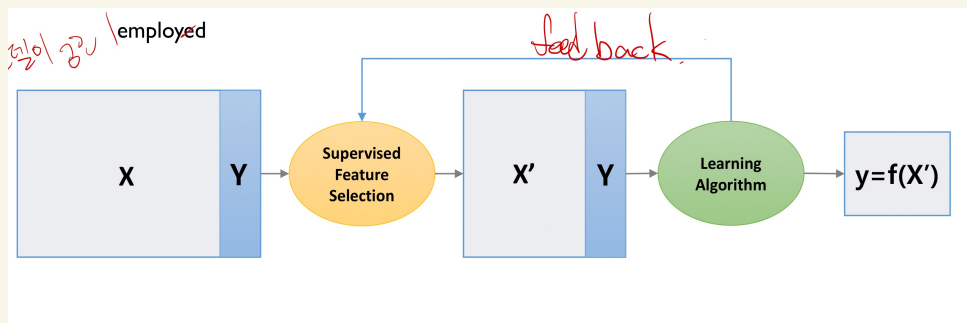
⇒ 2차원으로 축소해도 95%의 정보를 잃지 않음.
($16 \times 16 \rightarrow 2$)

- 일반적으로 도입된 변수가 증가할 수록 모델의 성능을 증가한다. (형식적으로는 이러함.)

- Supervised vs Unsupervised Method.

- Supervised.

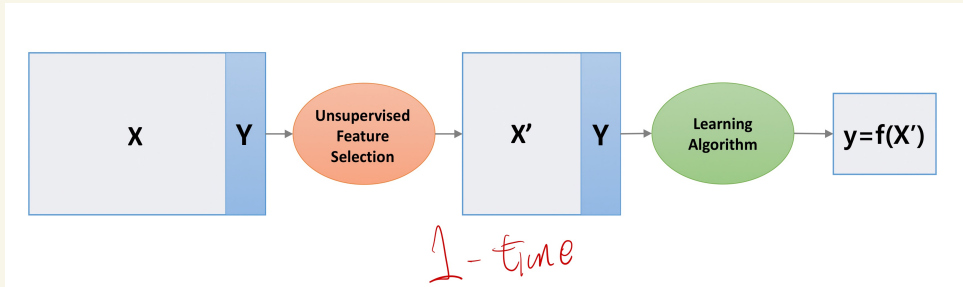
- 모델이 좋았는지 개선, (feed Back Loop 논리)



- Unsupervised

· 무감독 학습에 해당 X

· 1-time Learning



· feature selection vs extraction

