

# 5주차 복습

4조 부도현 임청수 노지예 한세림

# 목차 A table of contents.

---

## ● 경사하강법

1. 최소제곱법의 계산
2. 경사하강법
3. 경사하강법의 한계

## ● 특성공학

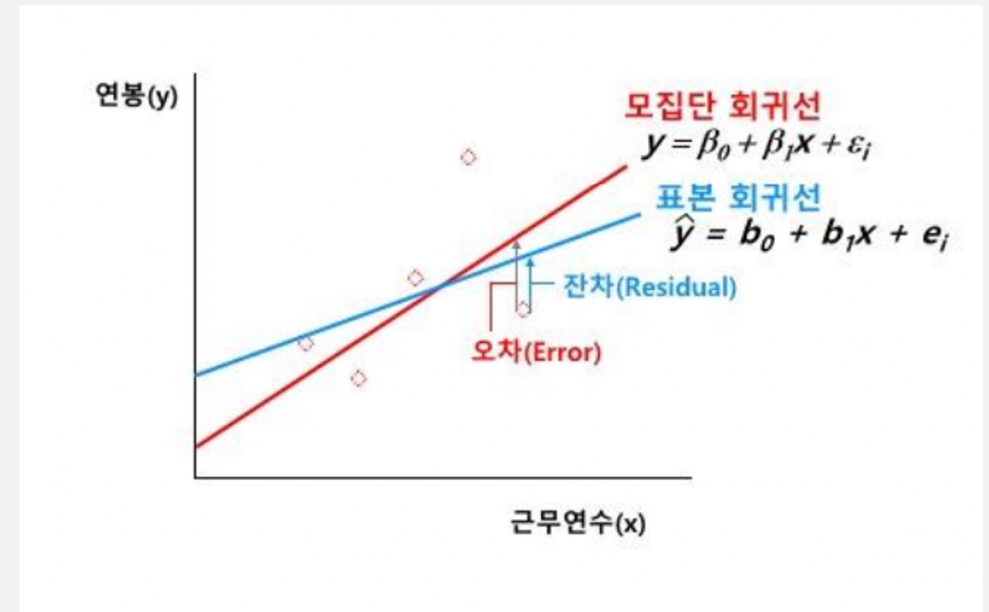
1. 특성공학의 개념
2. 특성공학의 방식



# PART 1 경사하강법

# ◆ 모델을 어떻게 평가할 수 있을까?

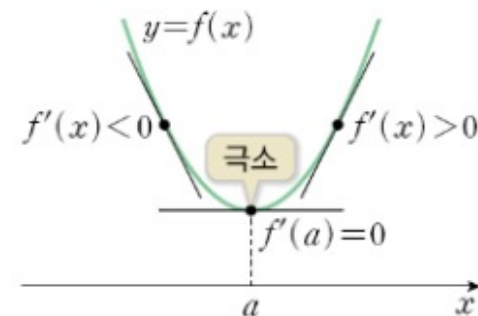
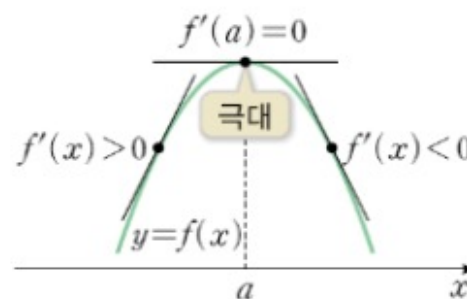
- 우리는 주어진 데이터에서 **데이터의 경향**을 가장 잘 설명하는 모델을 찾고 있다.
- 데이터의 경향을 가장 잘 설명한다는 것은 **실제 데이터와 예측값의 차이가 작음**을 의미한다.
- 실제 데이터와 예측값의 차이가 작은 모델을 좋은 모델이라 부른다.



# ◆ 좋은 모델을 어떻게 구할 수 있을까?

"틀린 정도"를 함수로 표현한 비용함수(cost function)의 최솟값을 구한다.

함수의 최솟값 구하기  
= 미분계수가 "0"인 지점!



최소제곱법의 W와 b에 대하여 편미분을  
진행하여 0이 되는 지점을 계산!

$$\frac{\partial}{\partial \hat{\beta}_0} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)$$

$$\frac{\partial}{\partial \hat{\beta}_1} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i$$

## ◆ 전통적 최소제곱법 계산의 한계



현실함수는 닫힌 형태가 아니거나, 함수 형태가 복잡해서(비선형 함수 등) 미분계수 계산이 어려운 경우가 많다



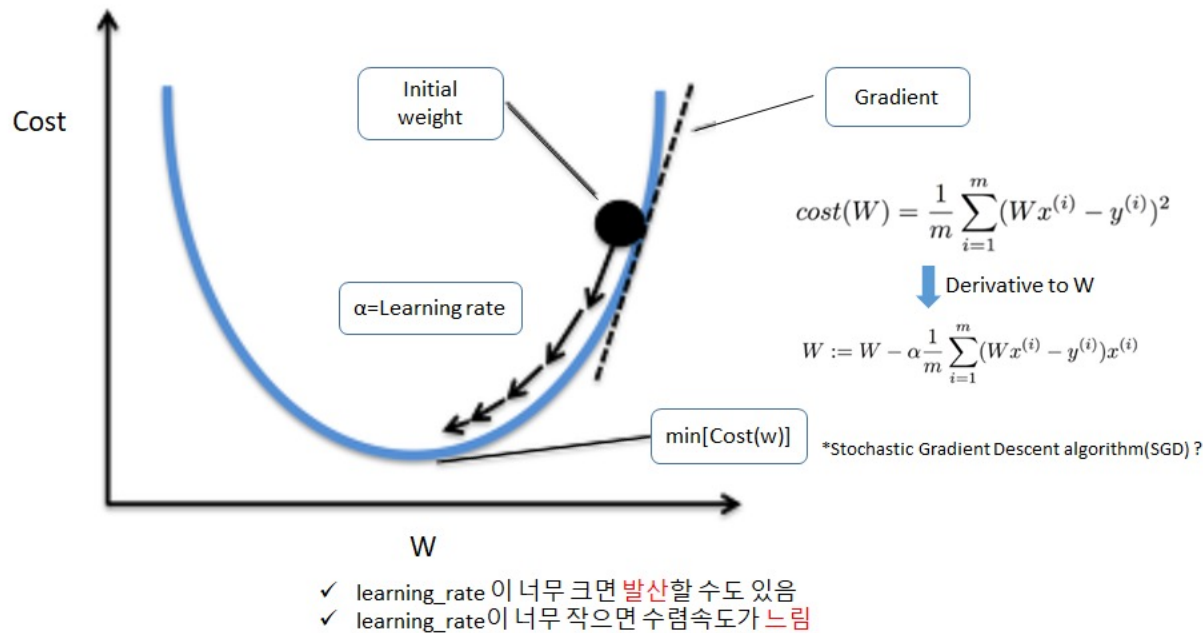
컴퓨터는 미분계수 계산보다 경사하강법을 구현하는 것이 더 편하다



데이터 양이 매우 큰 경우 경사하강법 알고리즘이 더 효율적이다



# ◆ 경사하강법이란



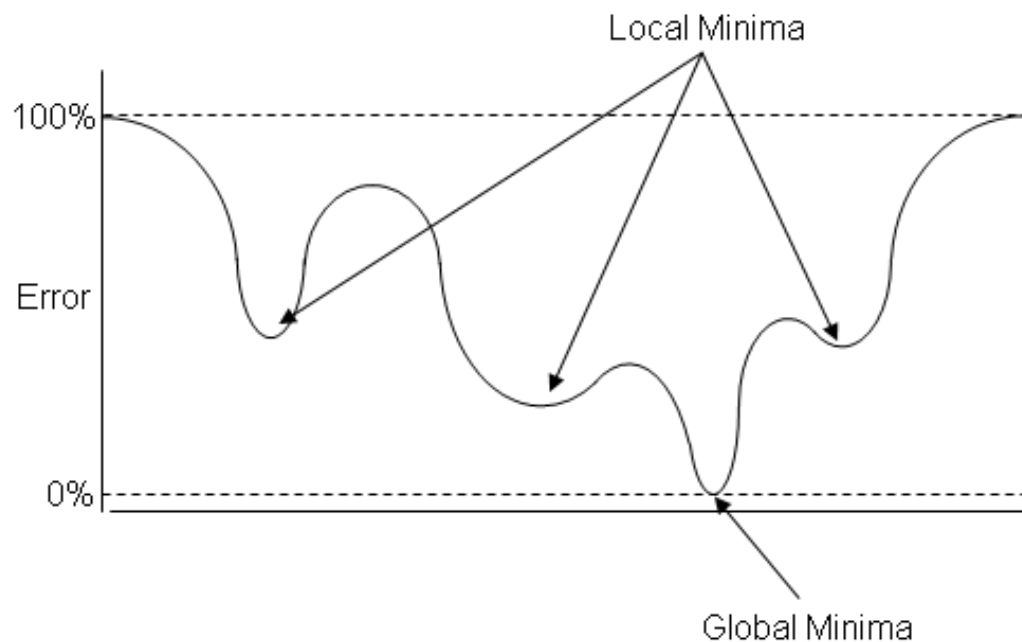
## ● 정의

함수의 기울기를 구한 후 기울기가 낮은 쪽으로 이동시켜 최솟값에 이를 때까지 반복하는 것이다.

## ● 대략적인 순서

1. 임의의 매개변수를 정해 비용함수의 x값으로 지정
2. 해당 매개변수로 모델의 오차를 구한 다음, 비용함수의 y값으로 지정
3. 시작 지점에서 다음 지점으로 갈 방향을 정하기 위해, 시작 지점의 기울기 계산
4. 기울기(Gradient)와 보폭(Learning rate)을 사용해 다음 지점으로 이동
5. 위 과정을 최솟값에 도달할 때까지 반복

## ◆ 경사하강법의 한계

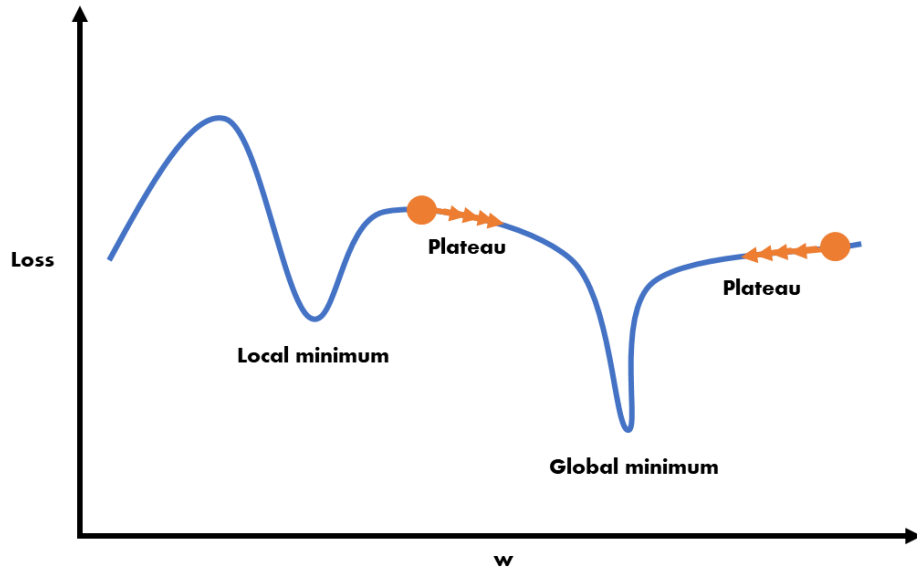


1. 데이터가 많아질수록 **계산량 증가**  
거대한 빅데이터를 사용하는 기계학습에 악영향을 준다

2. **Local minimum**(optima) 문제  
활성화 함수로 인해 그 모양이 울퉁불퉁해지고,  
그로 인해 최적해에 수렴하지 못할 수 있다



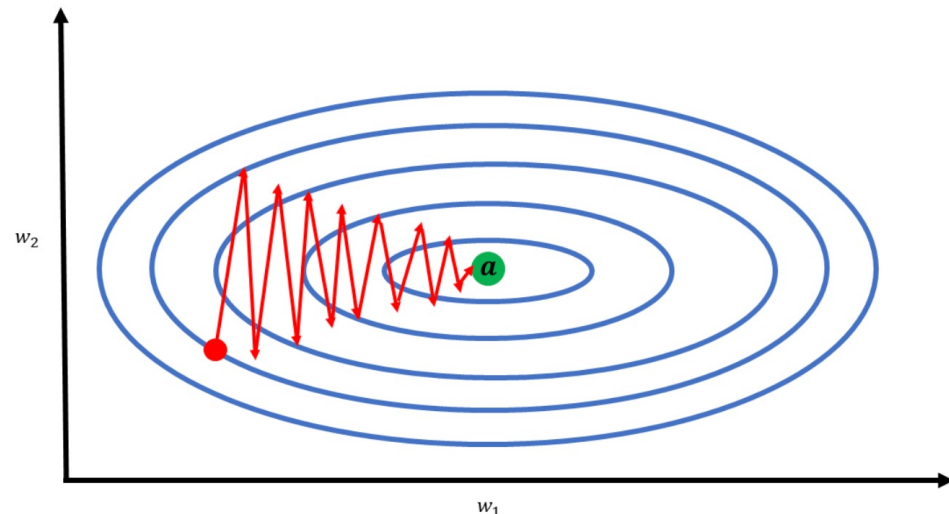
## ◆ 경사하강법의 한계



### 3. Plateau 문제

Plateau라 불리는 평탄한 영역에서는 학습속도가 느려지거나 정지할 수 있다

이로 인해 더 이상 학습이 일어나지 않는 **가중치 손실**(Gradient Vanishing) 현상이 발생할 수 있다

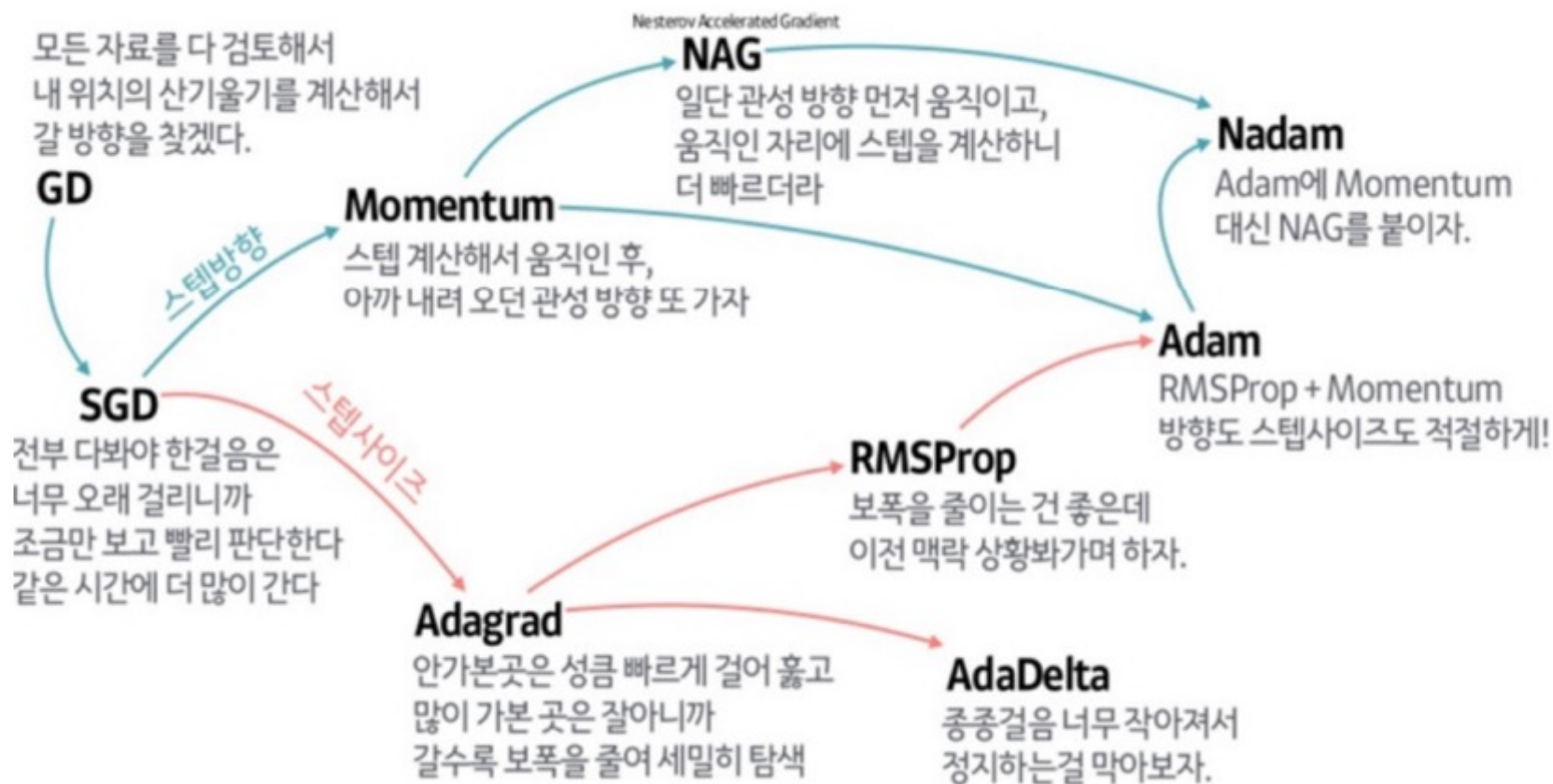


### 4. Zigzag 문제

찾아야 하는 해가 많아질수록 차원이 복잡해서 제대로 해를 찾지 못하는 문제이다

**가중치의 scale**이 다른 경우 발생한다

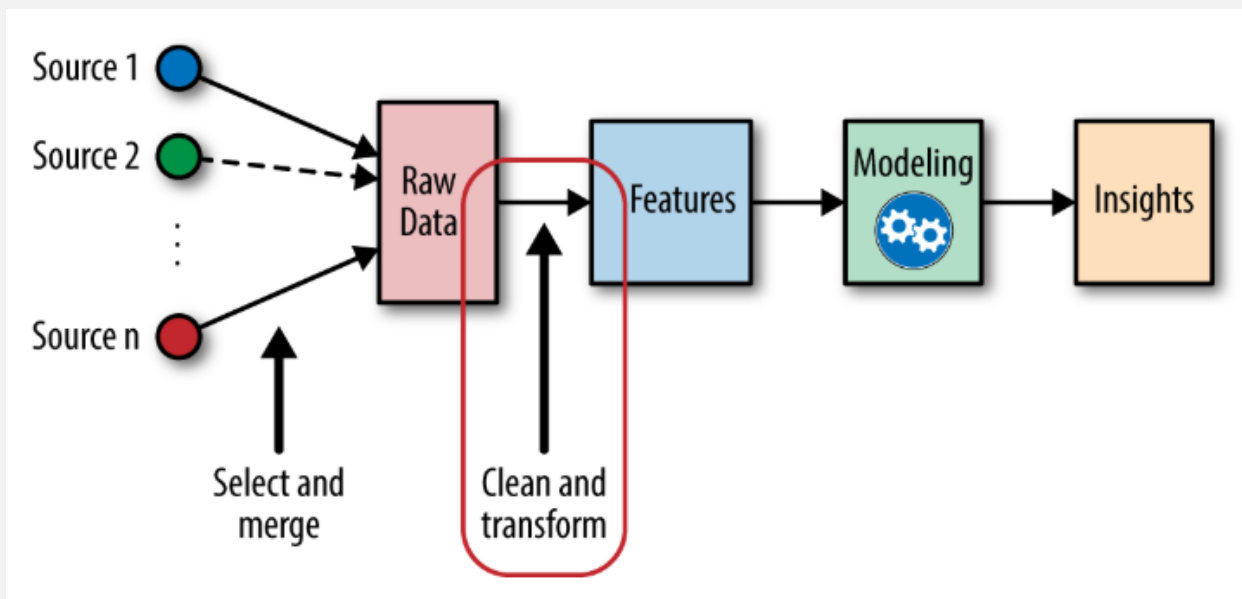
## ◆ 다양한 optimizer



# PART 2 특성공학

## ◆ 특성공학이란

- 특성공학은 **모델 훈련**에 사용하기에 가장 **적합**한 데이터 **특성**을 찾는 것
- Feature Engineering은 모델 성능에 미치는 영향이 크기 때문에 머신러닝 응용에 있어서 굉장히 중요한 단계이며, 전문성과 시간과 비용이 많이 드는 작업임



## ◆ 특성공학 방식

**특성 생성**

모델링 성능을 높이는 새로운  
특성을 만드는 과정

**특성 추출**

특성의 조합으로 유용한 특성을  
새롭게 생성하는 과정

**특성 선택**

관련없거나 중복되는 특성을  
필터링하는 과정

## ◆ 특성생성 기법

- 범주형 변수의 변환  
모델에 맞는 새 특성 생성

ex) 회귀모델을 위해 one hot encoding 적용

- 선형모델에서의 구간분할  
연속형 특성 > 특정 범위로 그룹화하여 이산형 특성 생성

ex) 시간 데이터를 오전, 오후로 그룹화

- 상호작용과 다항식  
X의 다항식을 새 특성으로 사용

ex) 특성 X에 대해 X의 제곱을 특성으로 추가

- 일변량 비선형 변환  
X에 log, exp, sin 함수를 적용해 새 특성생성

ex) log함수를 통해 데이터의 스케일 변경

# ◆ 차원이란

## ▶ 정의

데이터셋에서 변수, feature, 열(column)을 나타낸다.

## ▶ 차원이 많으면?

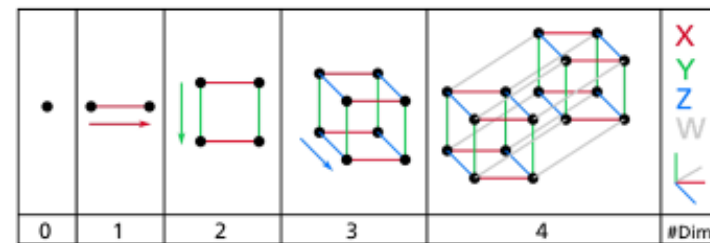
데이터의 특성(feature)은 데이터의 다양한 측면을 보여 줄 수 있기에 많으면 좋다고 착각한다.

## ▶ 단점 존재

하지만 이러한 **많은 특성**(feature)은 **훈련을 느리게** 하기도 하고, 때론 좋은 솔루션을 찾는데 **방해**하기도 한다.

## ▶ 차원의 저주

우리는 이를 **차원의 저주**(curse of dementionality)라고 한다.

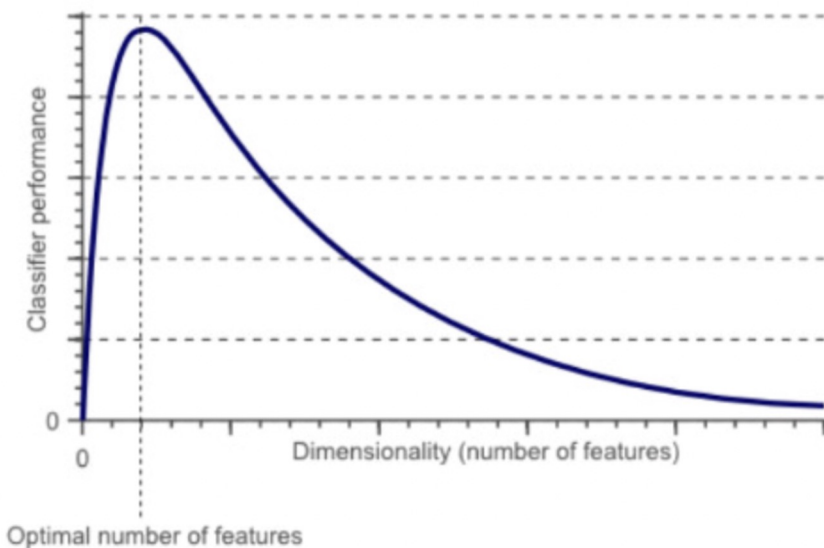


**Dimensionality Reduction**

**< Principal Component Analysis (PCA) >**



## ◆ 차원의 저주



출처: 구글(Google) 이미지, 차원이 커질수록 모델의 성능은 0에 수렴

- <차원의 저주 예시>
- 차원 (D=1) 성별 : 1000명 중 500명
- 차원 (D=2) 연령 :
- 30대 남성 1000명 중 200명
- 차원 (D=3) 거주지역 :
- 경기도 거주 30대 남성 1000명 중 50명

▶ 이러한 차원의 저주를 피하기 위해 **차원 축소**가 필요하다!

▶ 차원축소에는 특성 추출과 특성 선택이 있다.

## ◆ 특성 추출

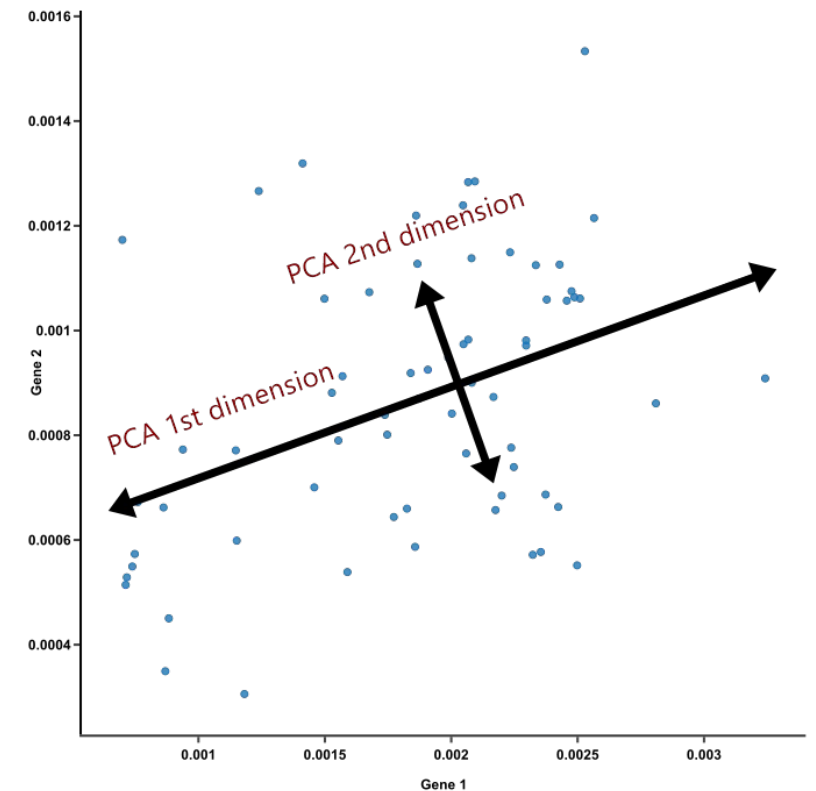
### ▶ 특성추출의 2가지 분류

Linear methods : PCA(Principal Component Analysis) ,  
LDA(Linear Discriminant Analysis) 등

Non linear methods : AE(AutoEncoder),t-SNE 등

### ▶ PCA 기법

PCA는 각 변수(Feature)를 하나의 축으로 투영시켰을 때 분산이 가장 큰 축을 첫번째 **주성분**으로 선택하고 그 다음 큰 축을 두번째 **주성분**으로 선택하고 데이터를 선형 변환하여 **다차원을 축소**하는 방법이다.



# 특성 선택

## 정의

- 특성선택은 모델을 구성하기 위한 특징을 선택하는 과정이다.
- 특성선택을 통해 모델의 복잡도를 감소시켜 모델의 성능을 향상시키고 처리 속도를 증가시킨다.

### <특성선택 전략>

**Filter based method** : information grain, chi-square test, fisher score, correlation coefficient, variance threshold

**Wrapper based bethod** : 전진선택, 후진 제거, 단계별 선택

**embedded method** : LASSO , RIDGE

**감사합니다**