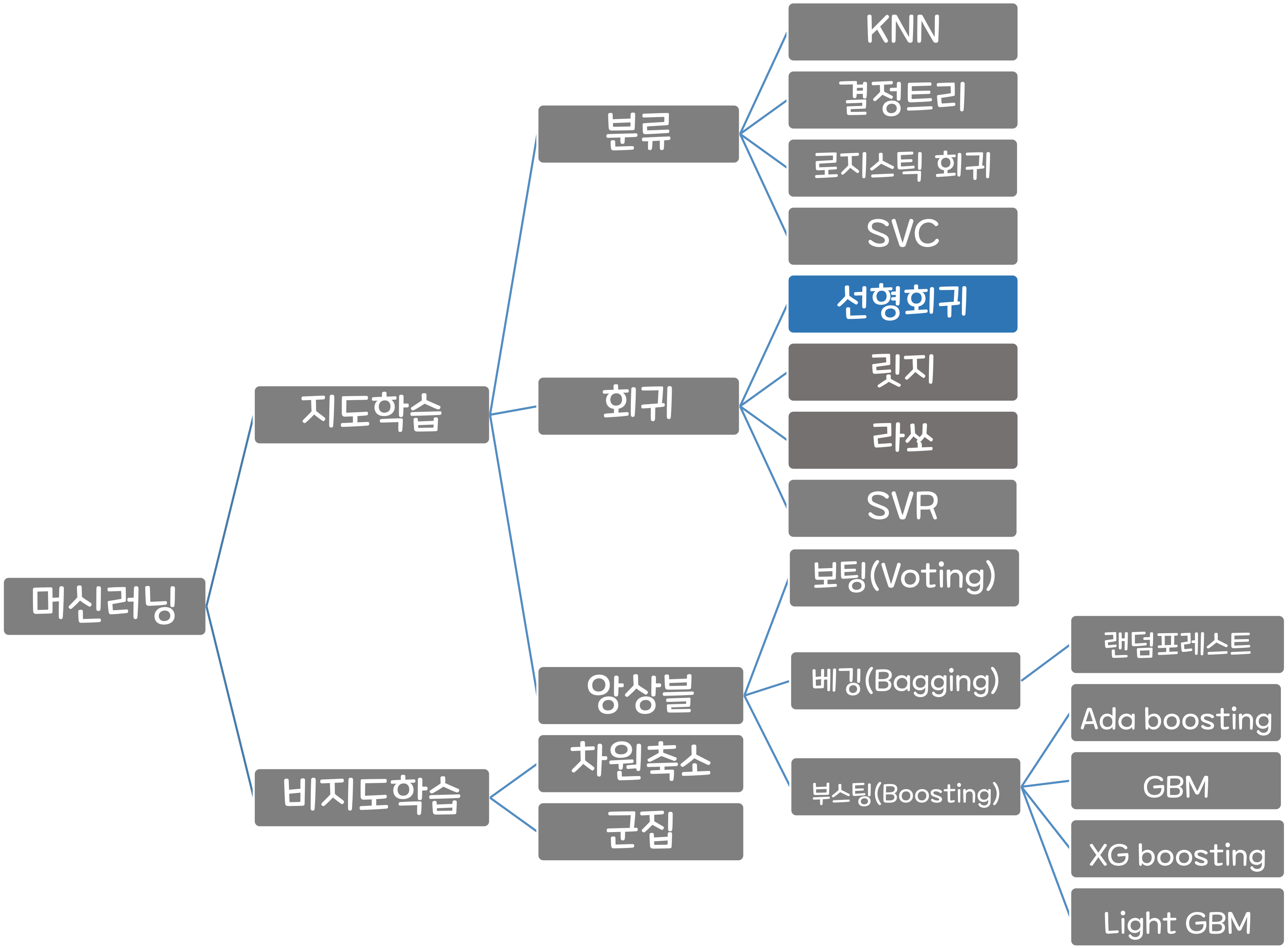
스마트인재개발원
Smart Human Resources Development

Machine Learning

chap 7. Linear Regression, L1(Lasso) 규제, L2(Ridge) 규제

김 민 수 연구원

머신러닝 개략도


머신러닝
지도학습
비지도학습
분류
회귀
앙상블
차원축소
군집
KNN
결정트리
로지스틱 회귀
SVC
선형회귀
릿지
라쏘
SVR
보팅(Voting)
베깅(Bagging)
부스팅(Boosting)
랜덤포레스트
Ada boosting
GBM
XG boosting
Light GBM

학습 목표

- 회귀 및 선형 회귀의 개념과 필요성을 이해하고 숙지한다.
- 회귀 모델의 평가지표를 이해하고 그 종류를 숙지한다.
- 경사하강법의 개념과 그 종류를 이해하고 숙지한다.

선형 회귀 수업 흐름도

개요
- 개념
- 활용분야

기본 함수식

$y = wx + b$

평가지표
- MAE
- MSE
- RMSE
- R2 Score

코드실습

모델의 목표
- 최적 w, b값 찾기 (경사하강법)

장 · 단점

규제 모델
- L1 : 라쏘(Lasso)
- L2 : 릿지(Ridge)

코드실습

선형 회귀(Linear Regression)란?

직선의 형태를 가지는 1차식으로 **연속적인**

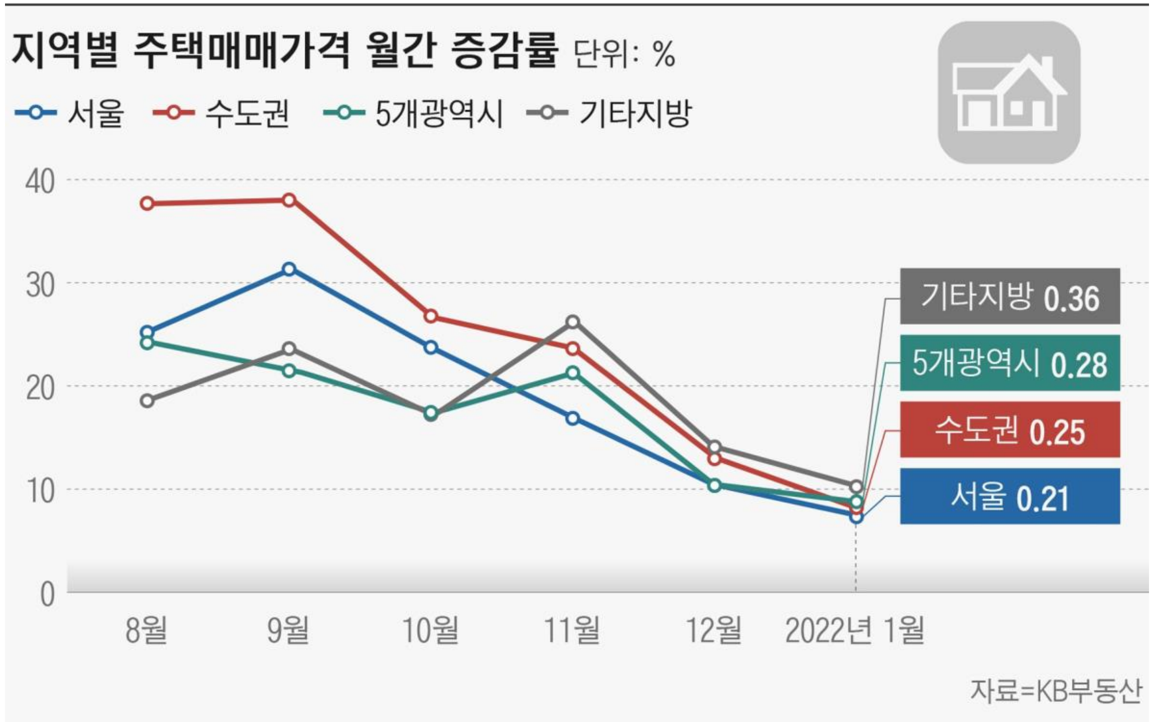
연속적인 값을 예측하는 분야

여러 개의 독립변수 x(특성)와 종속변수 y(예측값)의 선형 상관 관계를 모델링

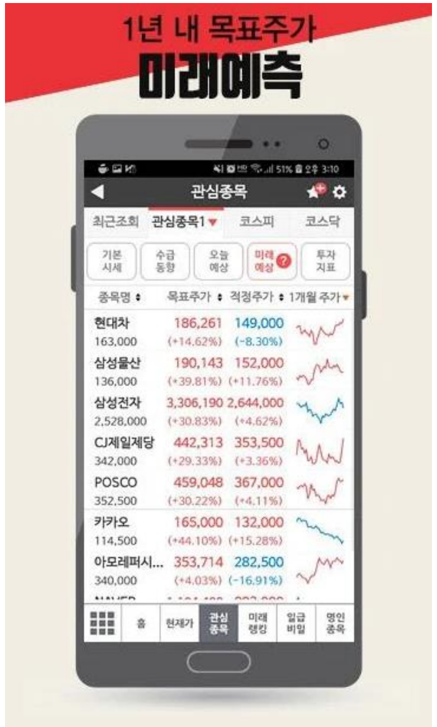
회귀의 중요성 및 필요성

스마트인재개발원
Smart Human Resources Development

- 선형 회귀는 규제가 있는 회귀 모델(Lasso, Ridge)과 딥 러닝 이론의 기초
- 회귀는 현업에서 많이 사용되며 활용 분야가 매우 넓음

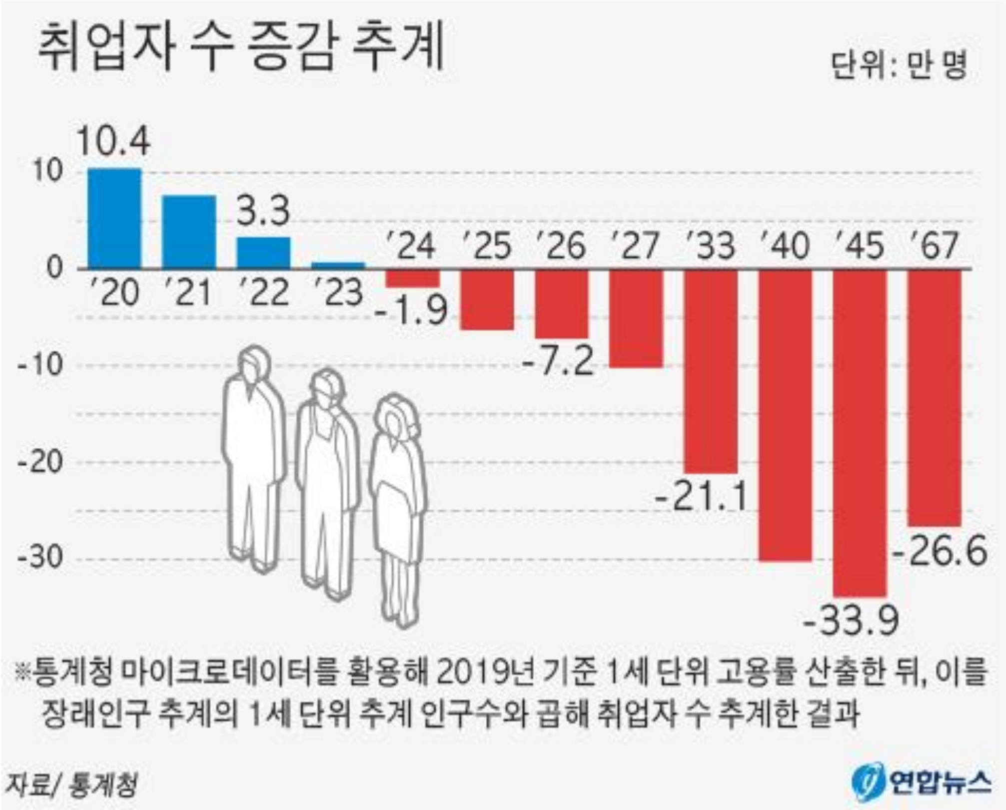


집값 예측

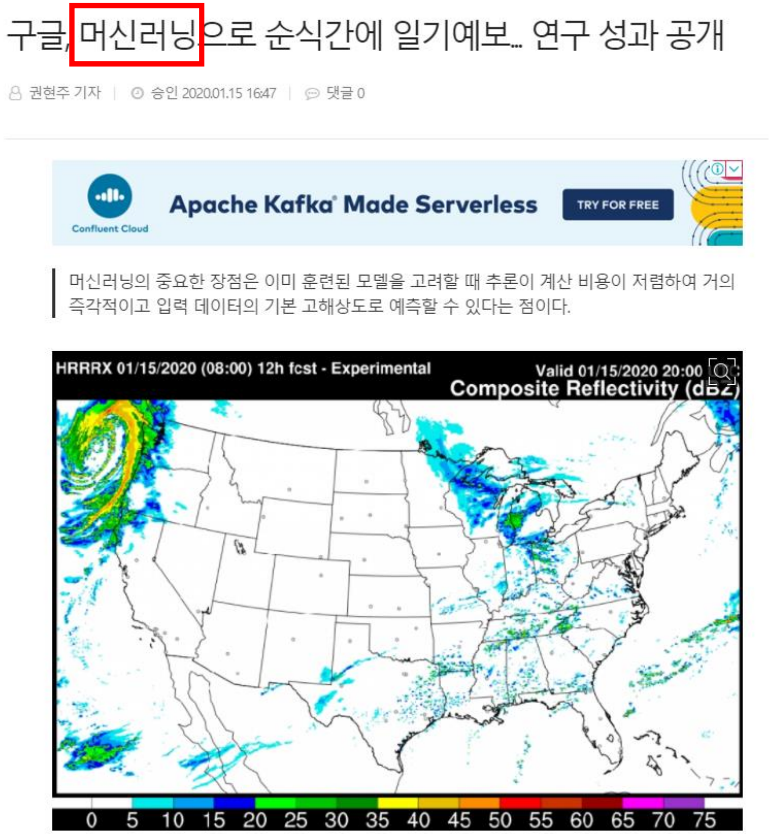


주가 예측

회귀의 중요성 및 필요성



통계 자료 예측



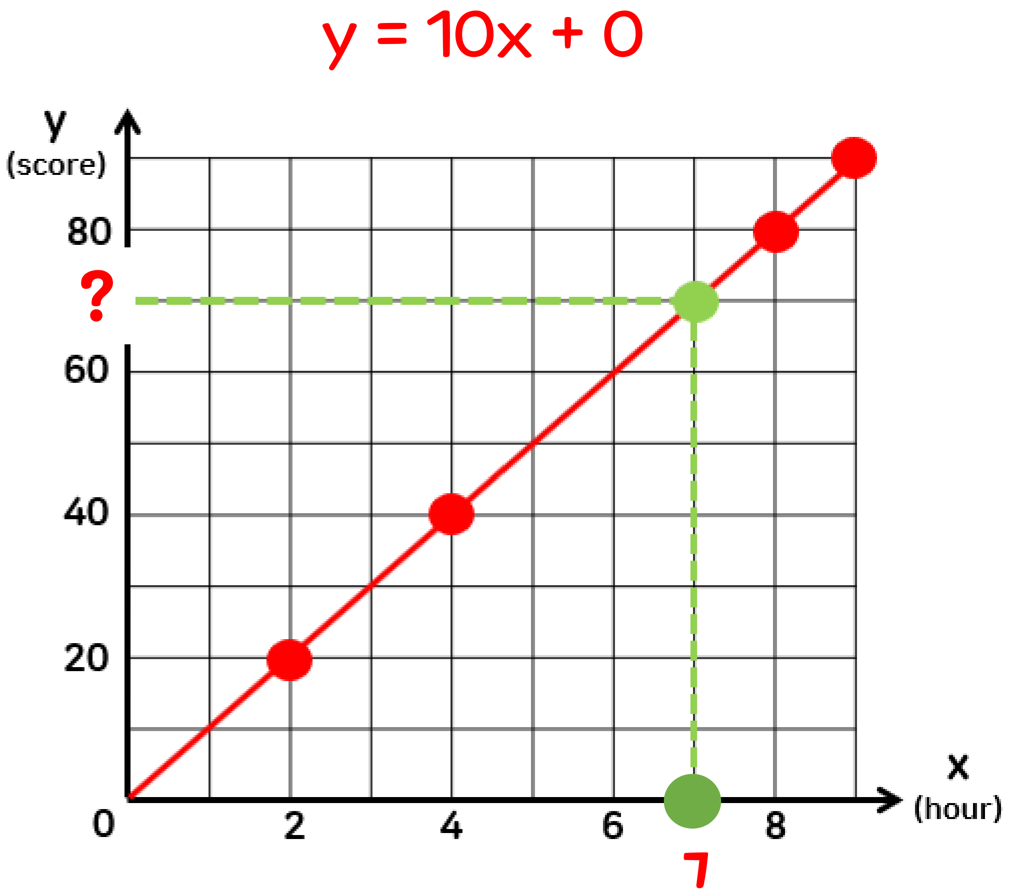
기상 예측

선형 회귀(Linear Regression)

x(hour)	y(score)
9	90
8	80
4	40
2	20

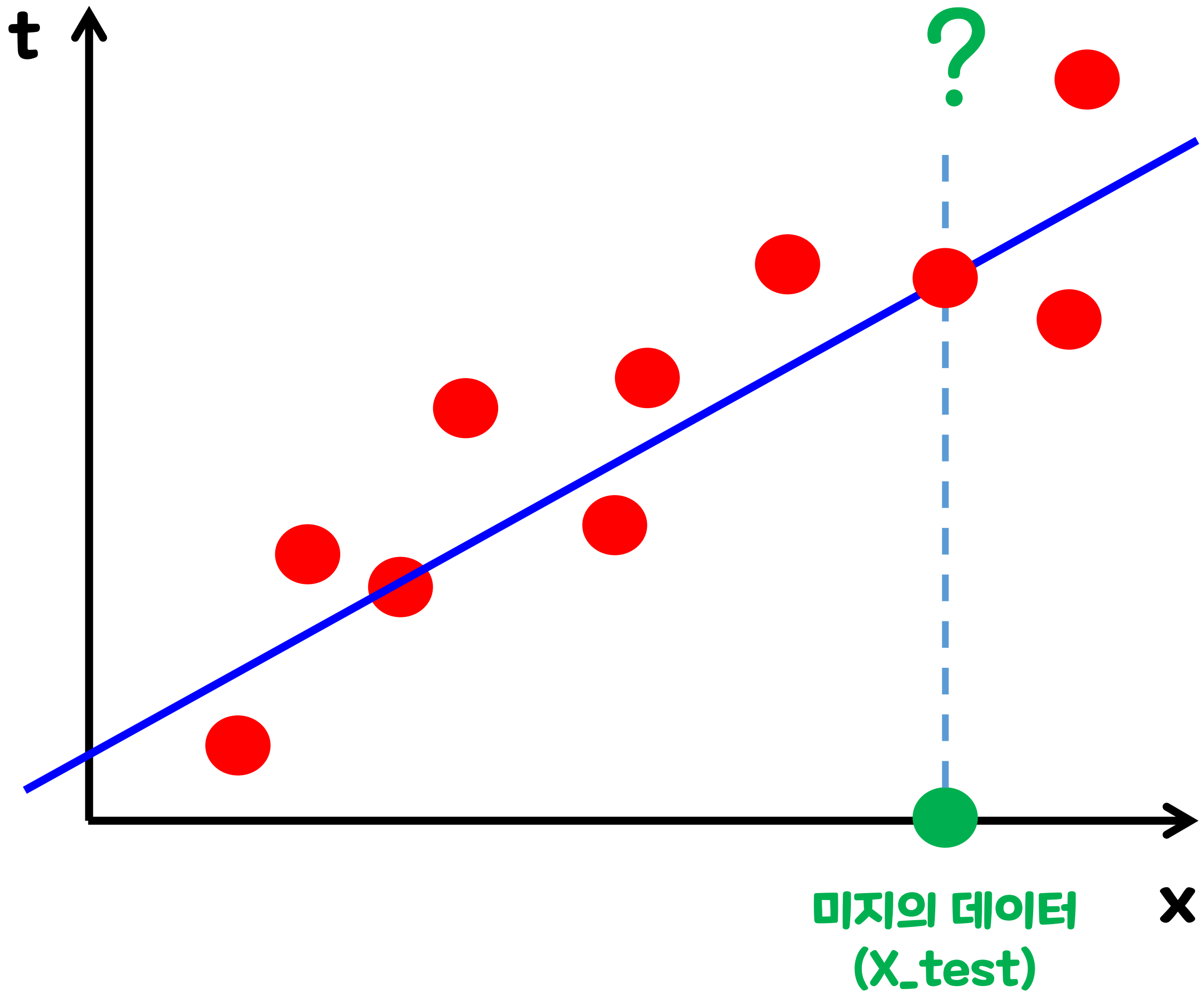
시험성적 데이터

7시간 공부 시 성적은?



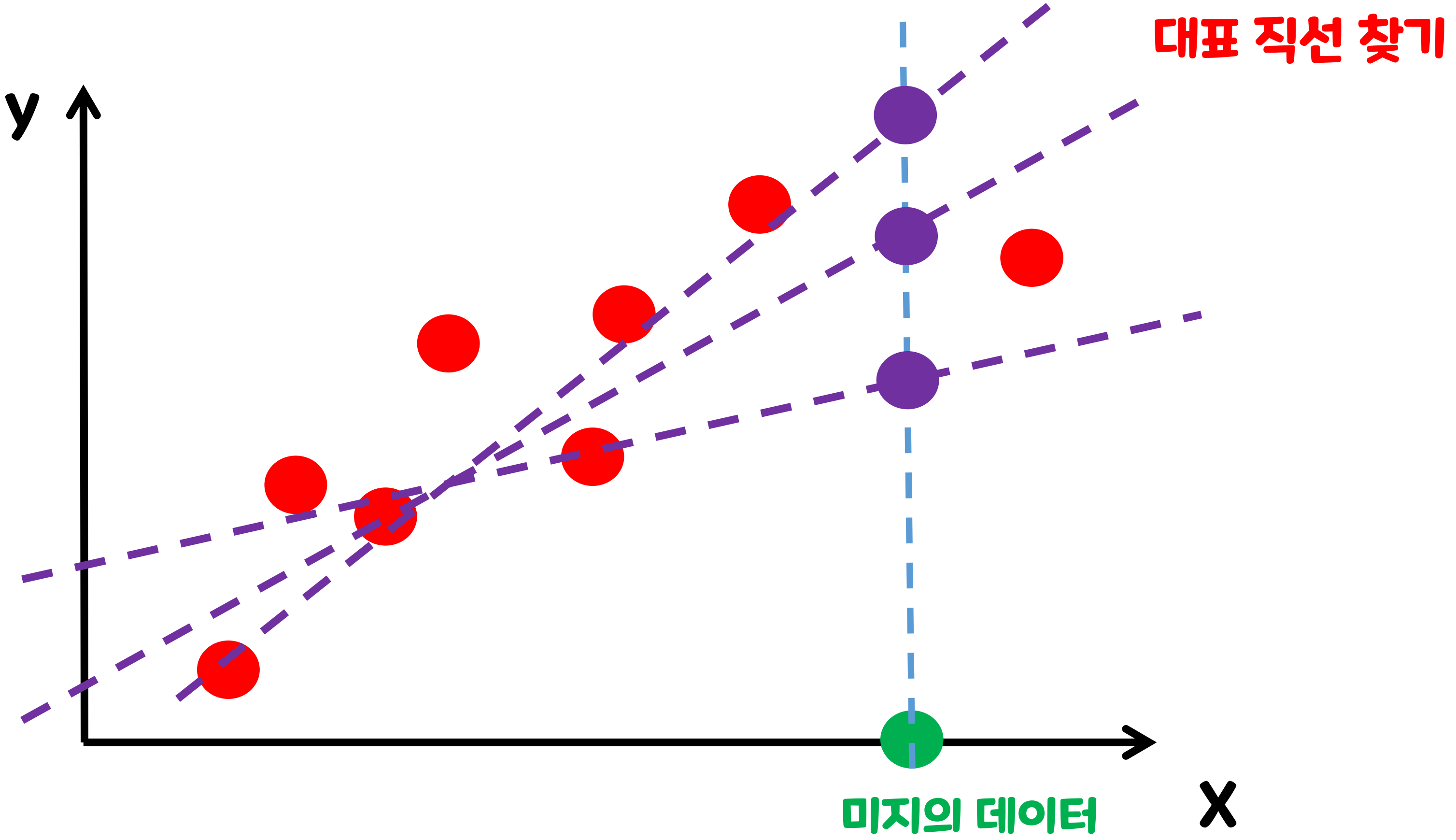
선형 회귀(Linear Regression)

X_train	y_train
공부시간(x)	시험성적(t)
9	74
14	81
21	86
27	88
32	90
37	92



선형 회귀(Linear Regression)

- 학습 데이터에는 없는 미지의 데이터에 대한 값을 예측할 때,
- 데이터의 분포를 가장 잘 표현할 수 있는 직선(y=wx+b)을 그려서 값을 예측하는 방법



선형 모델(Linear Model - Regression)

기울기(가중치)   절편(편향)

$$y = wx + b$$

종속(응답) 변수
=시험성적=예측값

독립(입력) 변수
=공부한 시간=특성값

다중 선형 회귀 함수

$$y = w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots + w_p x_p + b$$

- w : 가중치(weight), 계수(coefficient)
- b : 편향(bias), 절편(intercept)
-
- 모델 w 파라미터 : model.coef_
- 모델 b 파라미터 : model.intercept_

선형회귀 모델의 w, b 값 구하기 실습

회귀 모델의 성능은 어떻게 평가해야 할까?

MSE, RMSE, MAPE, R2 Score

회귀 모델 평가 지표

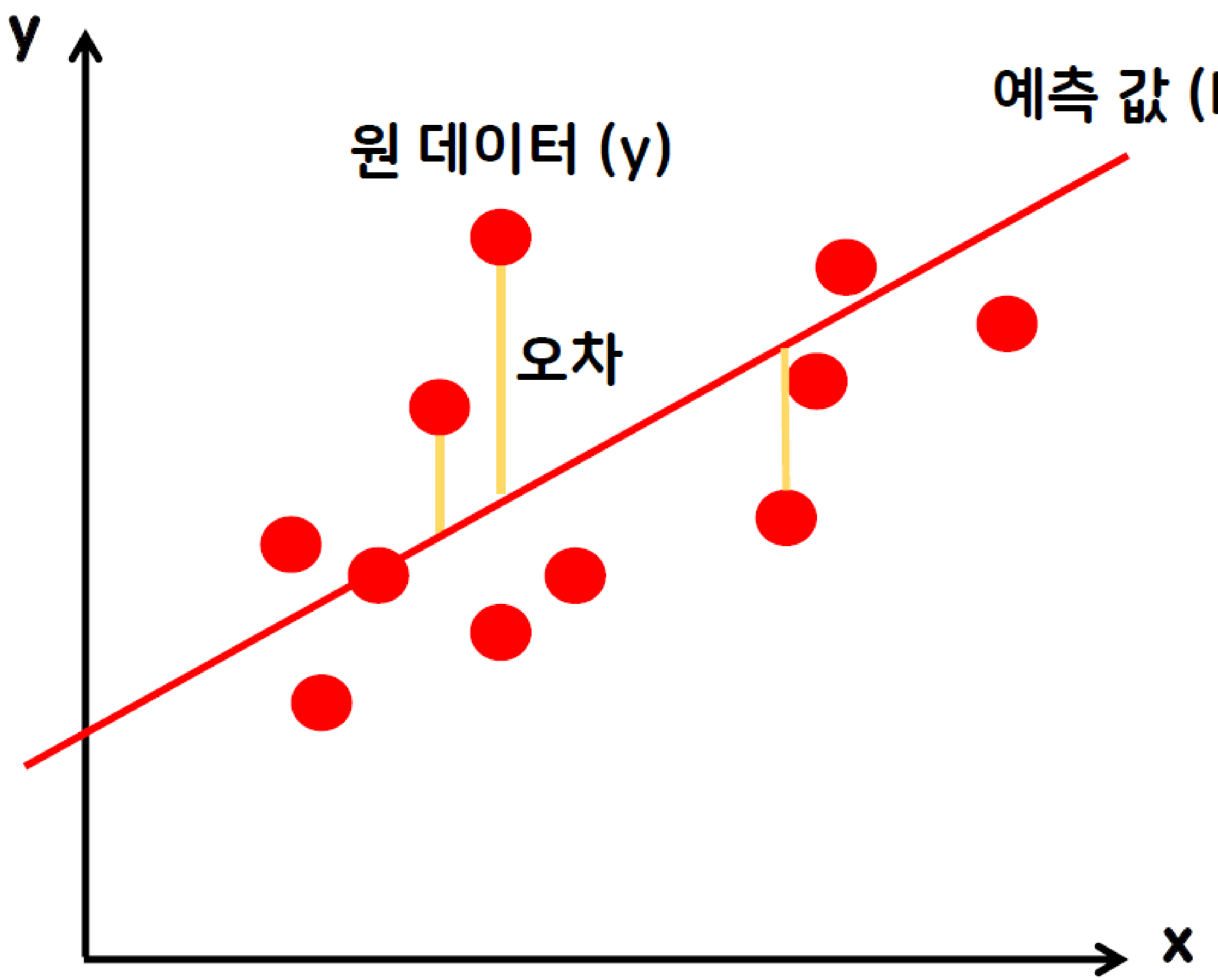
예측값

$$H(x) = w * x + b$$

예측값　정답값

$$\text{오차} = H(x) - y$$

절대값 또는
제곱 값 사용



평균제곱오차 (MSE : Mean Squared Error)

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\boxed{(\overset{\text{예측값}}{H(x_i)} - \underset{\text{정답값}}{y_i})}^2$$

오차

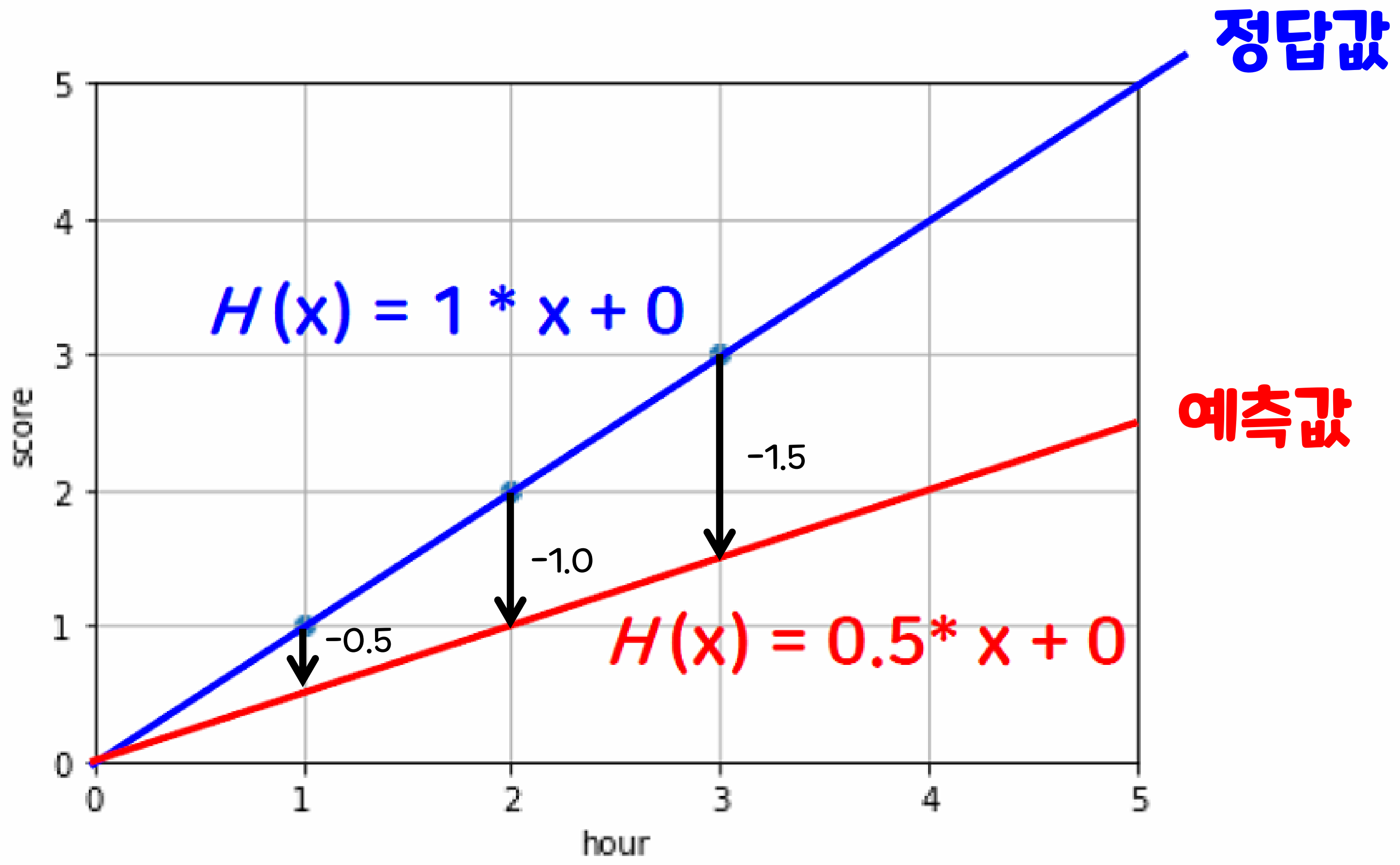
평균제곱근오차 (RMSE : Root Mean Squared Error)

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(H(x_i) - y_i)^2}$$

회귀 모델 평가 지표 종류

두 데이터의 MSE 값을 계산해보자.

x(hour)	y(score)
1	1
2	2
3	3



회귀 모델 평가 지표 종류

RMSE/MSE의 단점

- 예측 대상의 크기에 영향을 받음

삼성전자와 NAVER의 주가예측 RMSE가 5000이 나왔다면
두 모델의 성능은 동일한가요 ?

평균절대비율오차 (MAPE : Mean Absolute Percentage Error)

- 예측값과 실제값을 뺀(오차) 후 실제값으로 나눈 값의 평균 백분율로 표현하여 RMSE의 단점을 해결!

$$MAPE = \frac{100}{m}\sum_{i=1}^{m}\left|\frac{y_i - H(x_i)}{y_i}\right|$$

회귀 모델 평가 지표 종류

스마트인재개발원
Smart Human Resources Development

R2 Score

$$R^2 = 1 - \frac{\text{오차의 제곱의 합}}{\text{편차의 제곱의 합}}$$

- **회귀 함수(직선)가 평균에 비해 얼마나 그 데이터를 잘 설명할 수 있는가에 대한 점수**
- **편차** = 예측값과 평균과의 거리
- 오차 = 예측값과 회귀 직선과의 거리
- 일반적으로 0에서 1사이의 값이지만 예측이 심하게 어긋날 경우 – 값이 나올 수 있음
  (– 값이 나온다는 것은 회귀 직선이 평균보다 더 데이터를 잘 설명하지 못한다는 뜻)

회귀 모델의 최종 목표

MSE가 최소가 되는 최적의 w, b값 구하기

Linear 모델을 이용해 보스턴 집 값 데이터를 활용하여

주택 가격을 예측해 보자

feature_names

- CRIM : 지역별 범죄 발생률
- ZN : 25,000평방 피트를 초과하는 거주 지역의 비율
- IDUS : 비 상업 지역 넓이 비율
- CHAS : 찰스강에 대한 더미 변수(강의 경계에 위치한 경우는1, 아니면 0)
- NOX : 일산화질소 농도
- RM : 거주할 수 있는 방 개수
- AGE : 1940년 이전에 건축된 소유 주택의 비율
- DIS : 5개 주요 고용센터까지의 가중 거리
- RAD : 고속도로 접근 용이도
- TAX : 10,000달러 당 재산세율
- PTRATIO : 지역의 교사와 학생 수 비율
- B : 지역의 흑인 거주 비율
- LSTAT : 하위 계층의 비율

평균 제곱오차(MSE)가 최소가 되는 w와 b를 찾는 방법

1. **수학 공식을 이용한 해석적 방법** (Ordinary Least Squares)

2. **경사하강법** (Gradiendt Descent Algorithm)

Linear Model – Regression (MSE)

수학 공식을 이용한 해석적 방법 (Ordinary Least Squares)

$$a\sum x^2 + b\sum x = \sum xy$$
$$a\sum x + bn = \sum y$$

$$a = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \sum X \sum X}$$

$$b = \frac{n\sum X^2 \sum Y - \sum X \sum XY}{n\sum X^2 - \sum X \sum X}$$

X(hour)	y(score)
1	1
2	2
3	3

Linear Regression 모델 내부에 구현되어 있다.

경사하강법(Gradient Descent Algorithm)

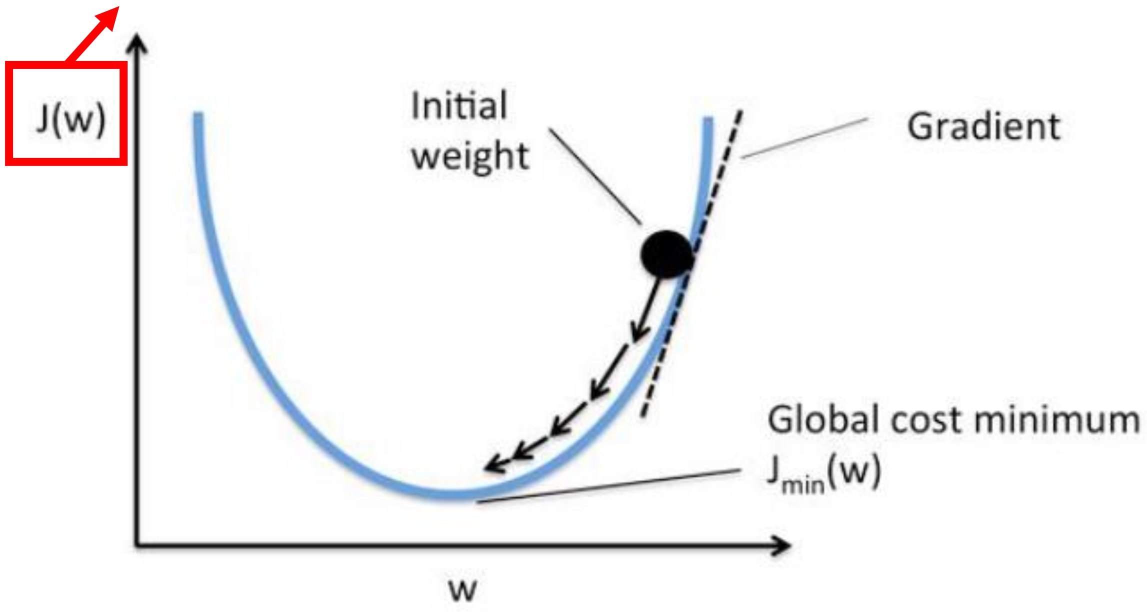
- 평균 제곱 오차(MSE)가 최소가 되게 하는 최적의 w, b값을 찾는 방법론
- 기계가 스스로 학습한다는 머신, 딥러닝의 개념을 있게 한 핵심 알고리즘

경사하강법 (Gradient Descent Algorithm)

경사하강법 (Gradient Descent Algorithm)

비용함수
= 오차
= MSE

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(H(x_i) - y_i)^2$$



**경사하강법
(Gradient Descent Algorithm)**

비용함수의 기울기(경사)를 구하여
기울기가 낮은 쪽으로 계속 이동하여
값을 최적화 시키는 방법

MSE가 최소가 되는 W와 b를 찾는 방법론

경사하강법 (Gradient Descent Algorithm)

우선 임의로 w값을 설정

(1) 최적의 w값을 찾아가기 위해서 시작점에서 손실 곡선의 기울기를 계산 → 비용함수를 w에 대해서 편미분
(2) 파라미터를 곱한 것을 초기 설정된 w값에서 빼 줌

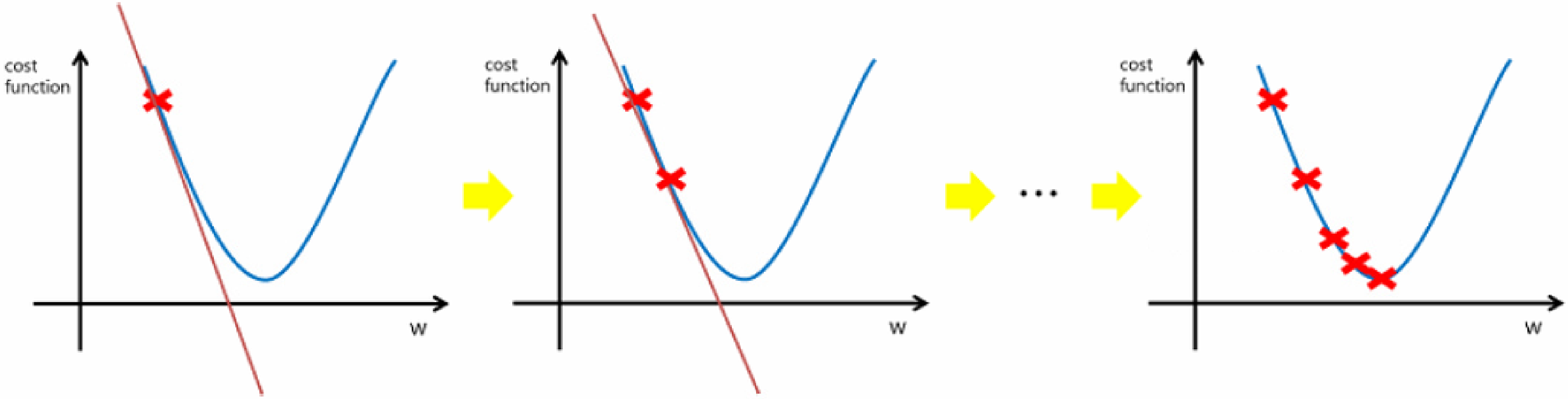
- 학습률(Learning rate) : 기울기의 보폭
- 학습률이 너무 작으면 최적의 w를 찾는데 오래 걸리고 크면 건너뛰어 버릴수 있음

$$w' = w - \eta \frac{\delta e}{\delta w}$$

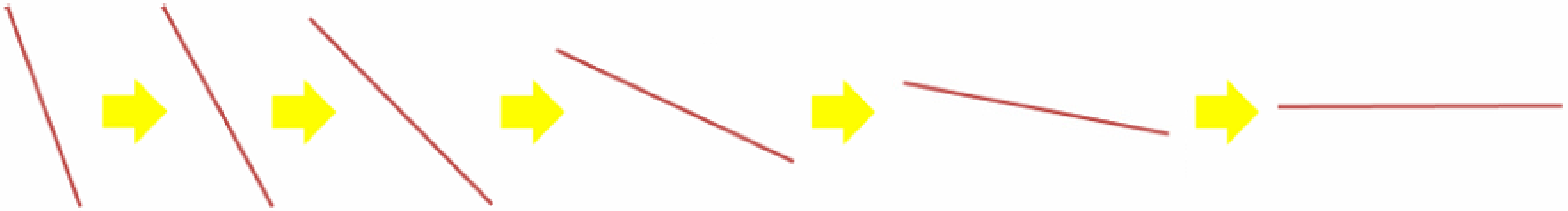
$$b' = b - \eta \frac{\delta e}{\delta b}$$

Linear Model - 경사하강법

경사하강법 (Gradient Descent Algorithm)



W값 갱신 과정

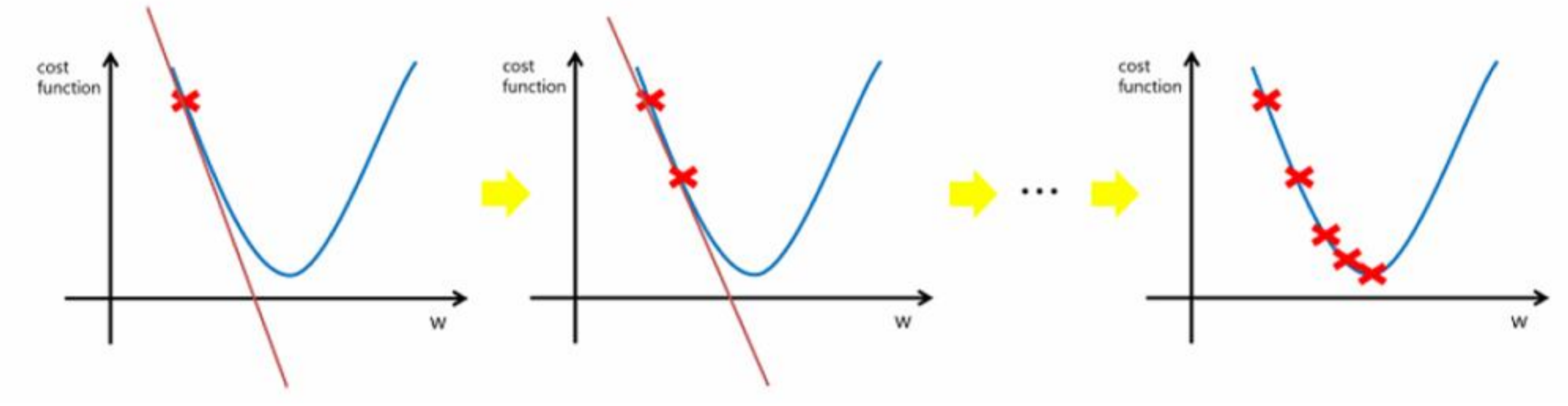


경사가 점차 감소되는 현상을 이용하므로 경사감소법!

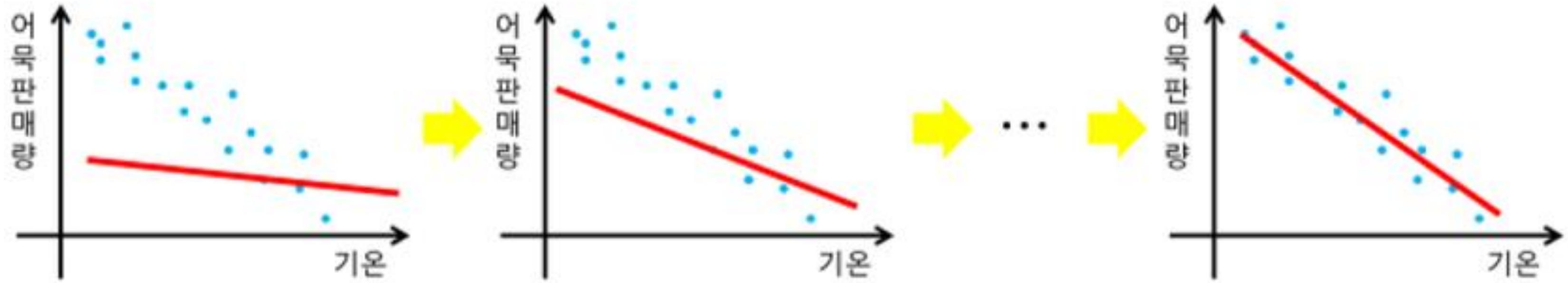
Linear Model - 경사하강법

경사하강법(Gradient Descent Algorithm)

최적의 w값을
찾는 과정



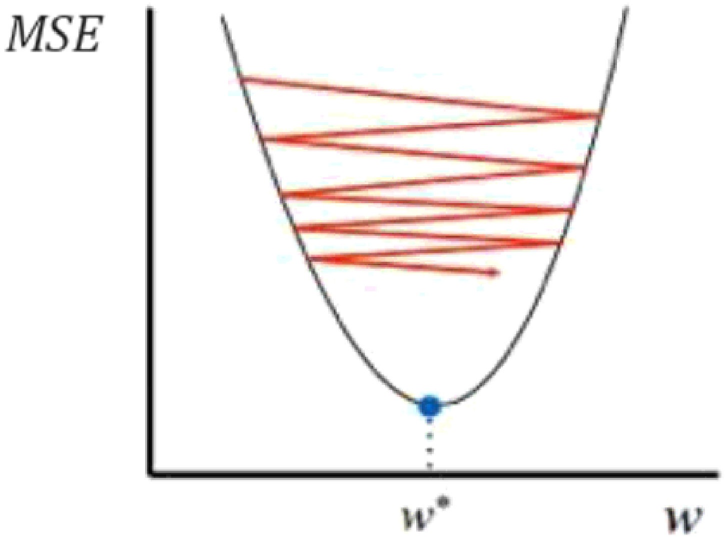
최적의 직선을
생성하는 과정



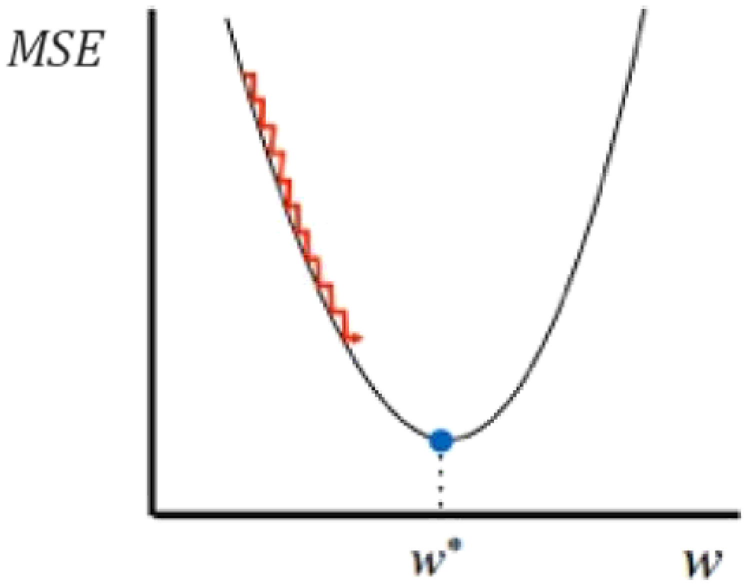
경사하강법 (Gradient Descent Algorithm)

학습률(Learning rate) : 기울기의 보폭

Learning rate가 큰 경우



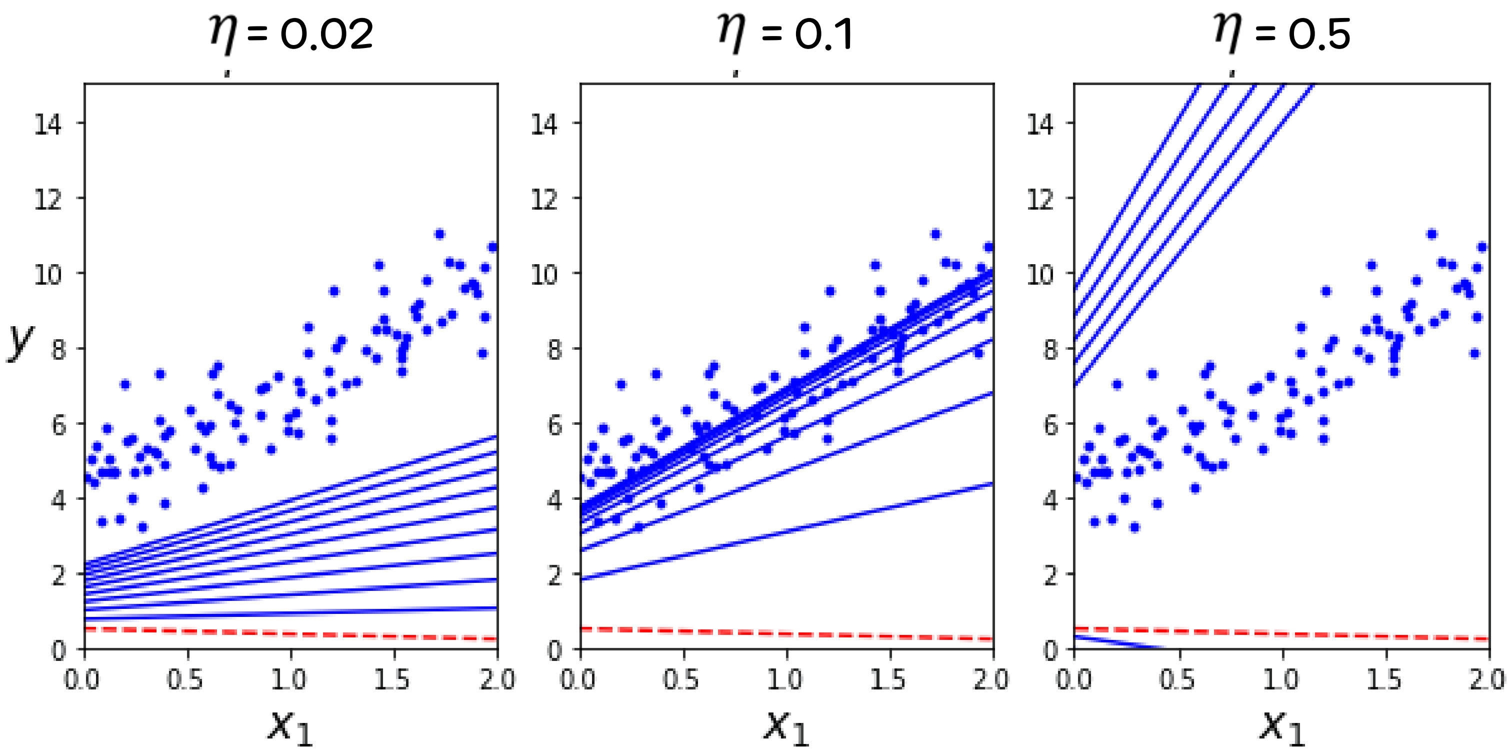
Learning rate가 작은 경우



학습률

$$w := w - \boxed{\eta} \frac{\partial}{\partial w} MSE$$

경사하강법 (Gradient Descent Algorithm)

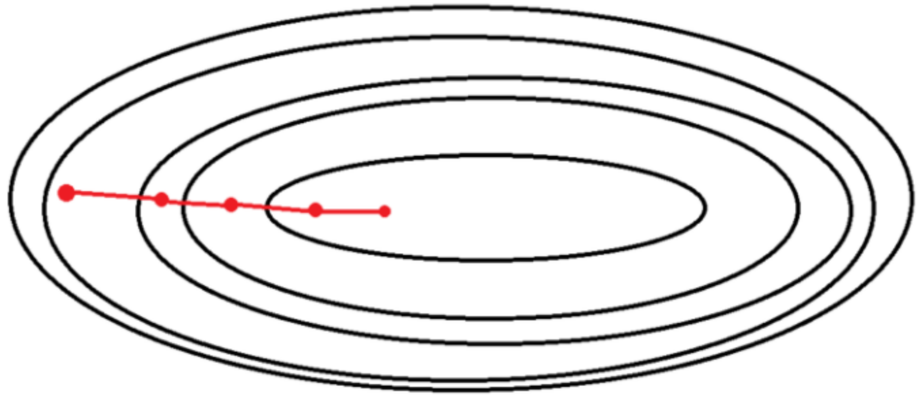


확률적 경사하강법(Stochastic Gradient Descent)

- 큰 데이터셋에서 일반 경사하강법의 느린 단점을 보완하기 위한 방식
- 전체 데이터가 아닌 일부 데이터만으로 w, b값을 업데이트
  - 따라서 항상 좋은 방향으로만 업데이트가 일어나지는 않음
  - 일부 데이터로 판단하기 때문에 속도가 빠름(데이터 수가 많을 때 유리)

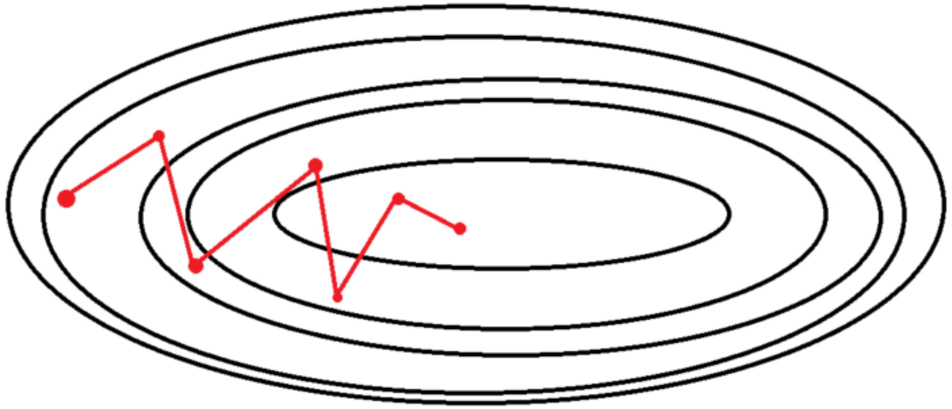
Linear Model – 확률적 경사하강법

스마트인재개발원
Smart Human Resources Development



경사하강법 (Gradient Descent Algorithm)

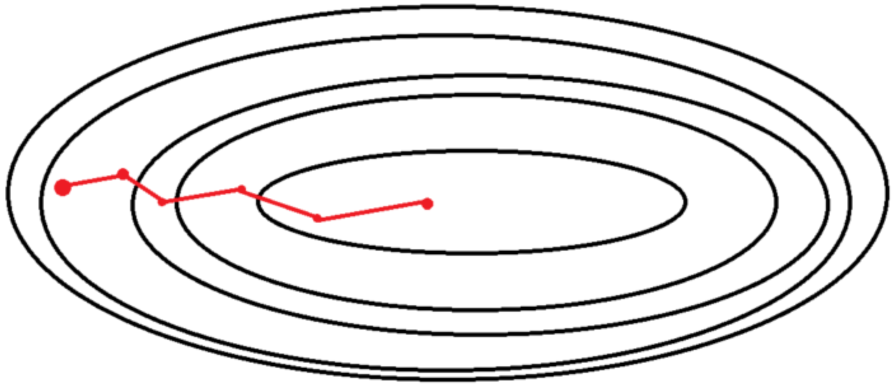
- 전체 데이터를 이용하여 경사를 구하기 때문에 최저점 수렴이 안정적
- 전체 데이터를 모두 한 번에 처리하기 때문에 속도가 느리고 메모리가 많이 필요

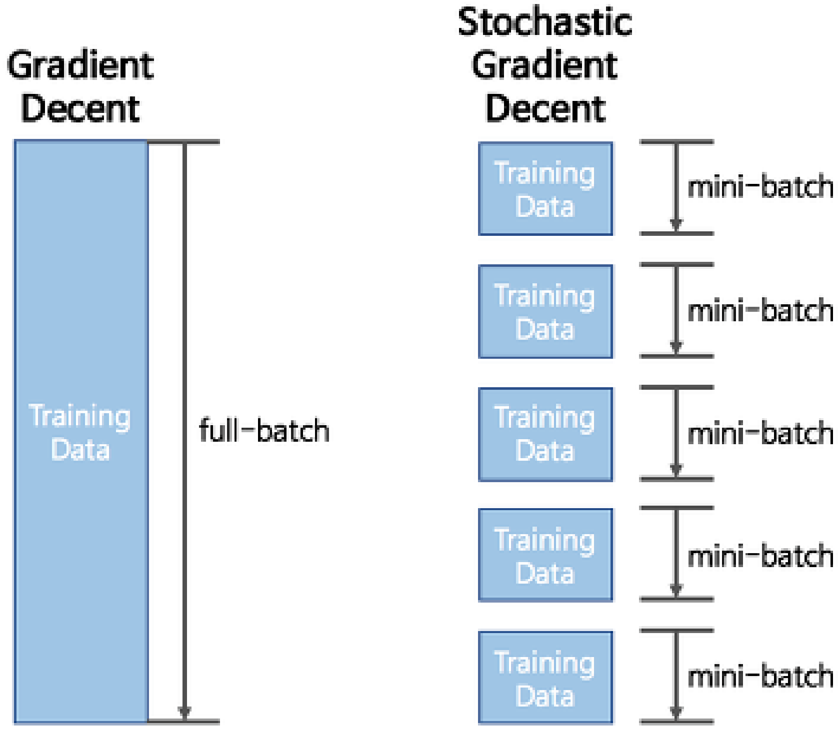


확률적 경사하강법 (Stochastic Gradient Descent)

- 전체 데이터 중 랜덤하게 선택된 하나의 **데이터**를 **이용하여 경사하강법을 진행**
- 적은 데이터로 학습할 수 있고 최적화 속도가 빠름
- 하나의 데이터를 이용하기 때문에 기울기의 방향이 크게 바뀌고(오차율↑) 최저점 안착이 비교적 힘듦

Linear Model – 미니배치 확률적 경사하강법





미니 배치 확률적 경사하강법
(Mini-Batch Stochastic Gradient Descent)

- 경사하강법과 확률적 경사하강법의 절충안
- 전체 데이터를 batch_size개씩 나눠 학습

Linear Model 장점

- 결과예측(추론) 속도가 빠르다.
- 대용량 데이터에도 충분히 활용 가능하다.
- 특성이 많은 데이터 세트라면 훌륭한 성능을 낼 수 있다.

Linear Model 단점

- 특성이 적은 저차원 데이터에서는 다른 모델의 일반화 성능이 더 좋을 수 있다. ➡ 특성확장을 하기도 한다.
- LinearRegression Model은 복잡도를 제어할 방법이 없어 과대적합 되기 쉽다.

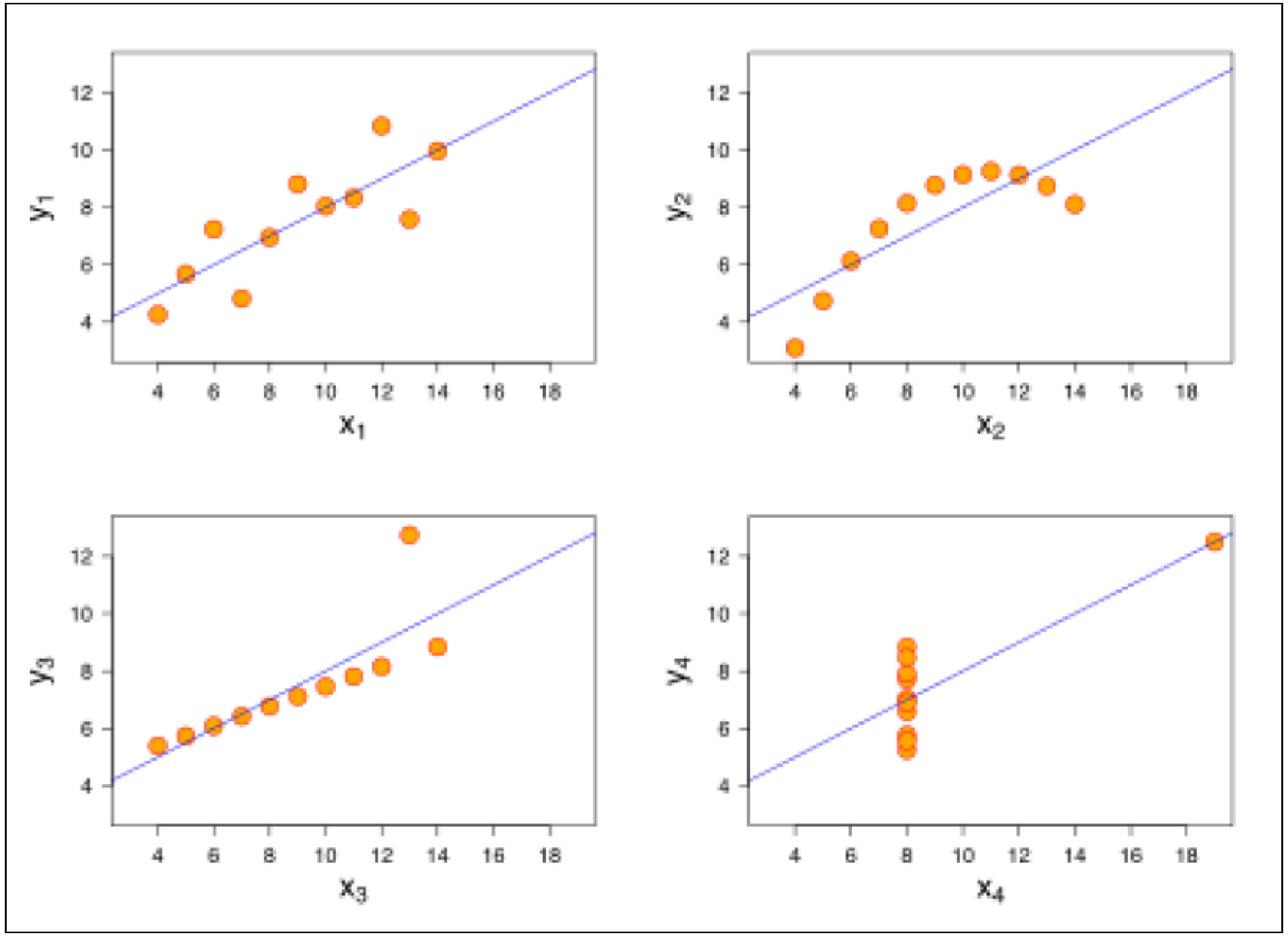
⬇

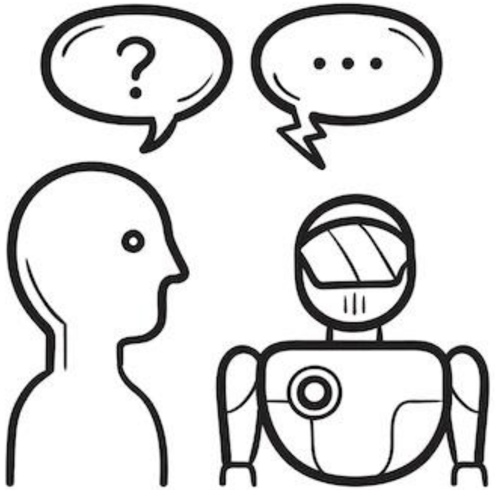
모델 정규화(Regularization)을 통해 과대적합을 제어한다.

선형 회귀(Linear Regression)

스마트인재개발원
Smart Human Resources Development

Linear Model 단점





규 제(Regularization)

L1 규제(Lasso), L2 규제(Ridge)

학습 목표

- 규제(정규화)의 개념과 필요성을 이해하고 숙지한다.
- 규제를 갖는 선형회귀 모델의 종류를 이해하고 숙지한다.

머신러닝 모델 개략도

스마트인재개발원 Smart Human Resources Development

- 머신러닝
  - 지도학습
    - 분류
      - KNN
      - 결정트리
      - 로지스틱 회귀
      - SVC
    - 회귀
      - 선형회귀
      - 릿지
      - 라쏘
      - SVR
    - 앙상블
      - 보팅(Voting)
      - 베깅(Bagging)
        - 랜덤포레스트
      - 부스팅(Boosting)
        - Ada boosting
        - GBM
        - XG boosting
        - Light GBM
  - 비지도학습
    - 차원축소
    - 군집

선형 회귀 수업 흐름도

스마트인재개발원
Smart Human Resources Development

개요
- 개념
- 활용분야

기본 함수식
- y = wx + b

평가지표
- MAE
- MSE
- RMSE
- R2 Score

코드실습

모델의 목표
- 최적 w, b값 찾기 (경사하강법)

장 · 단점

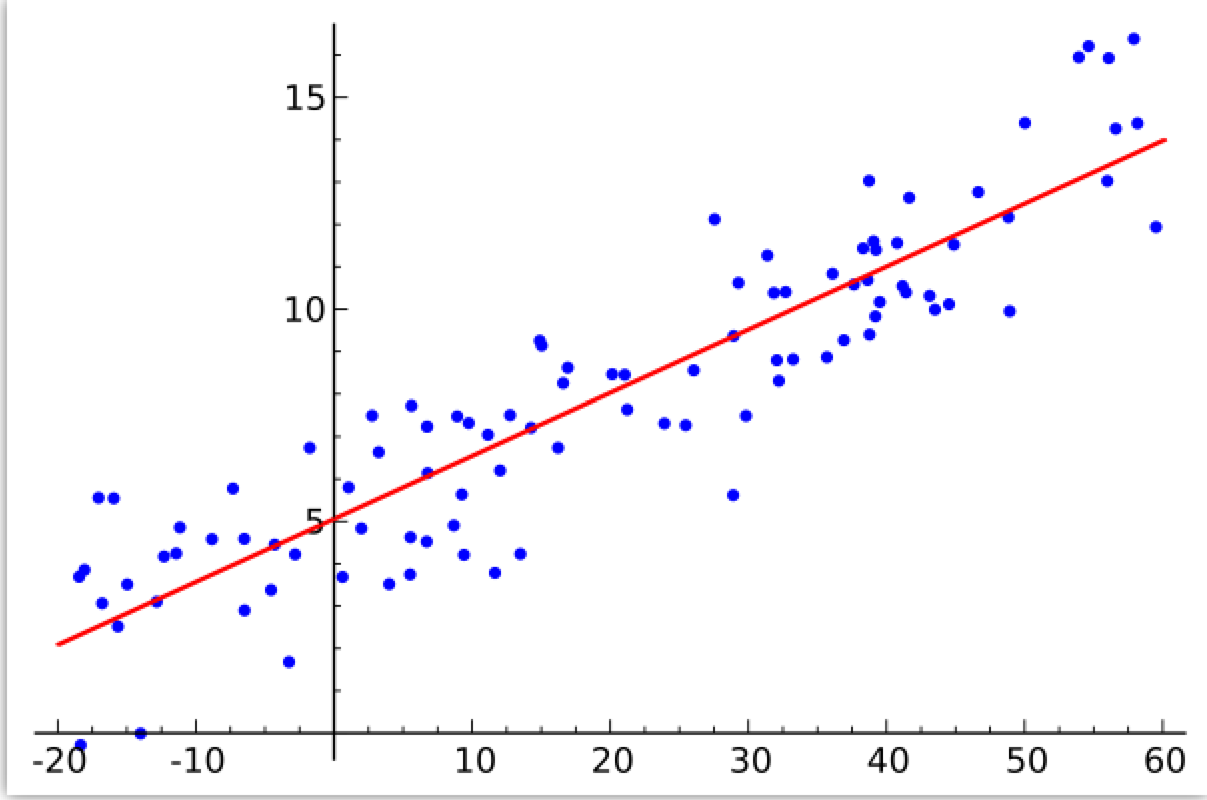
규제 모델
- L1 : 라쏘(Lasso)
- L2 : 릿지(Ridge)

코드실습

규제(Regularization)란?

선형 회귀 모델에서 과대적합의 위험을
감소시키기 위해 w값의 비중을 줄이는 것

선형 회귀(Linear Regression)



선형회귀 모델은 **학습 데이터를 전부 반영**하여 하나의 직선 방정식을 만들게 됨 →

학습 데이터에 과대적합 되는 것을 방지할 수 있는 방법이 없음

선형 회귀 - 모델 정규화(Model Regularztion)

스마트인재개발원
Smart Human Resources Development

모델 정규화

$$y = w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_px_p + b$$

- **w(회귀 계수)값이 크다 = 입력에 따른 예측 결과가 크게 바뀜 → 새로운 데이터가 들어오면 제대로 예측하지 못할 수 있음 = 과대적합의 위험이 높음**
- **w값을 적절히 낮게 조절하여 과대적합의 위험을 줄이는 것이 규제의 핵심**
- L1 규제 : Lasso

  w의 모든 원소에 똑같은 힘으로 규제를 적용하는 방법. 특정 계수들은 0이 됨.
  특성선택(Feature Selection)이 자동으로 이루어진다.

- L2 규제 : Ridge

  w의 모든 원소에 골고루 규제를 적용하여 0에 가깝게 만든다.

선형 회귀 - 모델 정규화(Model Regularztion)

스마트인재개발원 Smart Human Resources Development

규제 모델의 비용함수(Cost function)

규제의 강도(alpha)

비용함수

· L1 규제 : Lasso

$$\boxed{J(\theta)} = \mathrm{MSE}(\theta) + \boxed{\boxed{\alpha}\sum_{i=1}^{n}|\theta_i|}$$

L1규제 항

· L2 규제 : Ridge

$$J(\theta) = \mathrm{MSE}(\theta) + \boxed{\boxed{\alpha}\frac{1}{2}\sum_{i=1}^{n}\theta_i^2}$$

L2규제 항

· 선형회귀

$$J(\theta) = \mathrm{MSE}(\theta)$$

선형 회귀 - 모델 정규화(Model Regularztion)

정규화 : cost 함수

선형회귀에서 w 공식

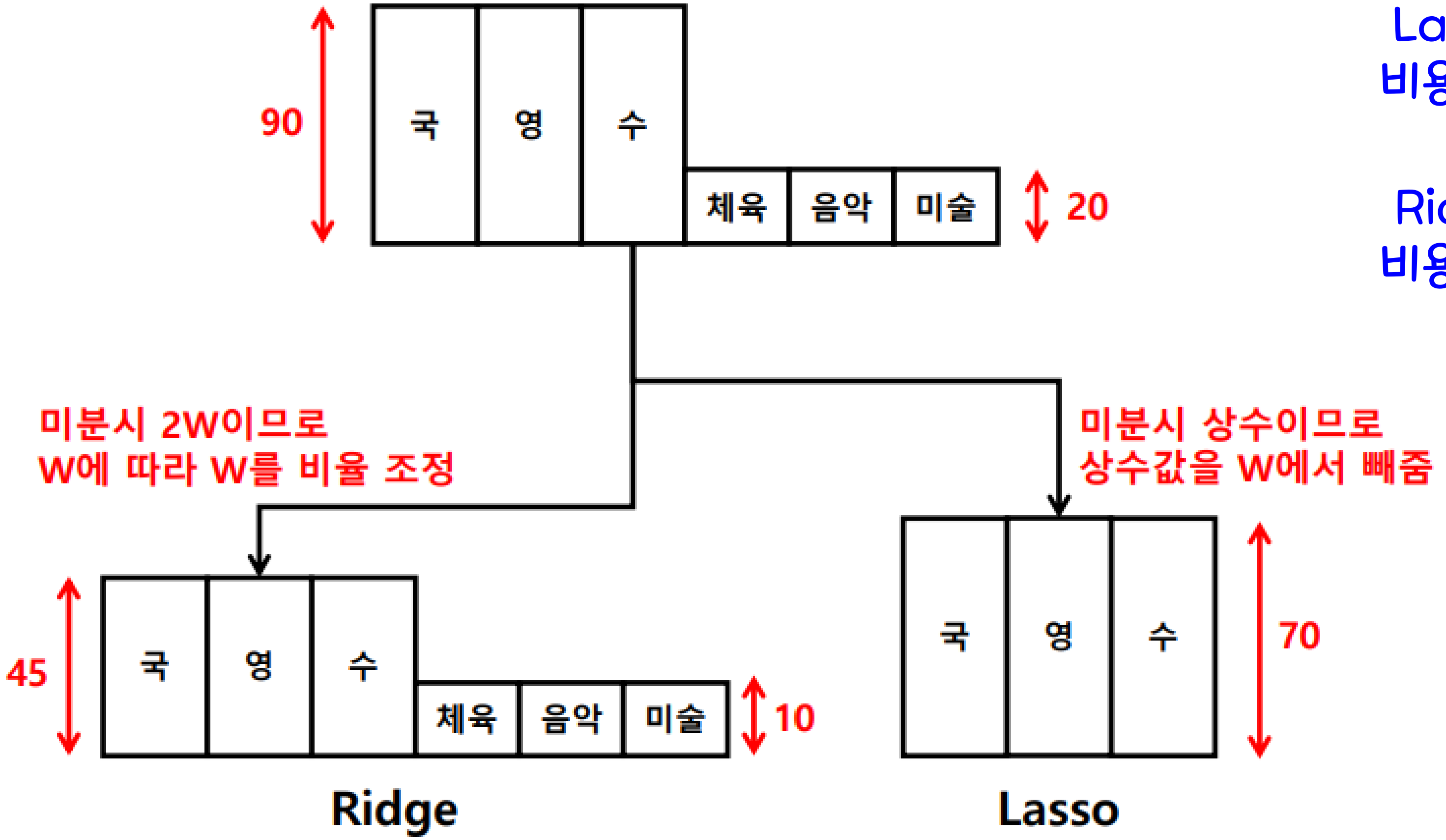
$$w := w - \eta \frac{\partial}{\partial w} MSE$$

Lasso 비용함수

$$\mathrm{MSE}(\theta) + \alpha \sum_{i=1}^{n} |\theta_i|$$

Ridge 비용함수

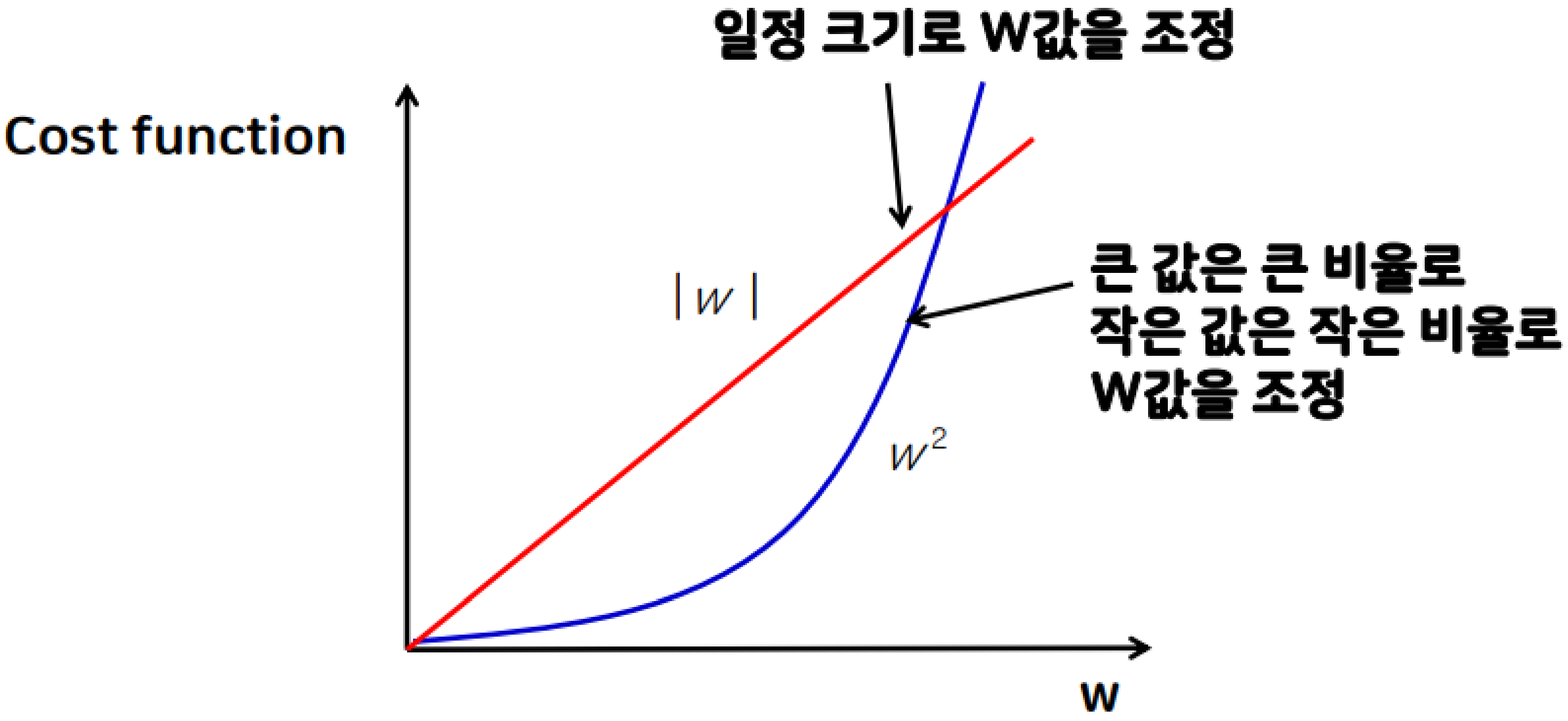
$$\mathrm{MSE}(\theta) + \alpha \frac{1}{2} \sum_{i=1}^{n} \theta_i^2$$



선형 회귀 - 모델 정규화(Model Regularztion)

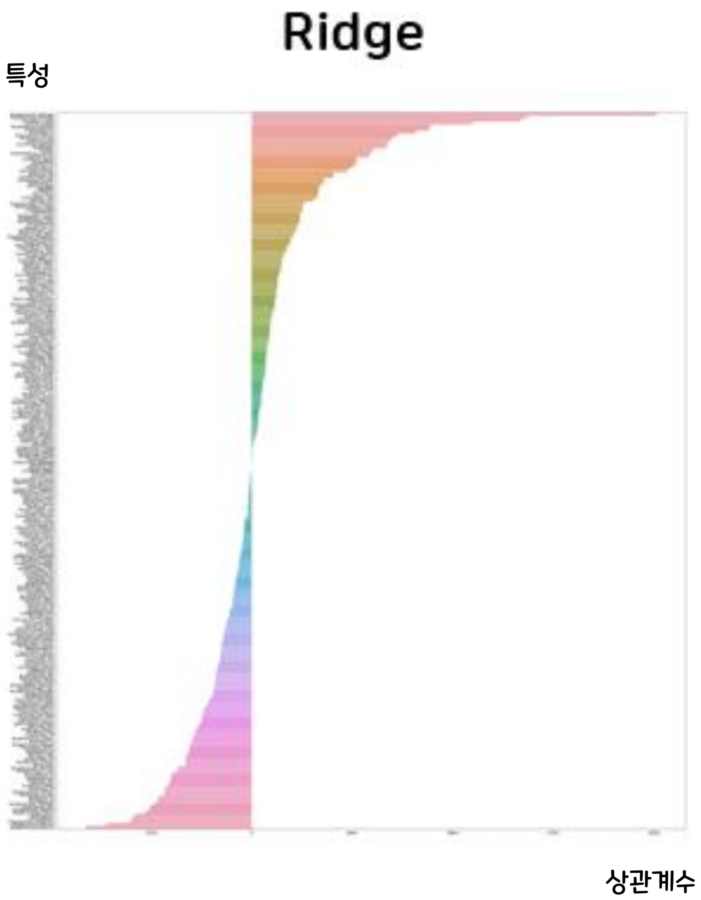
스마트인재개발원
Smart Human Resources Development

정규화 : cost 함수

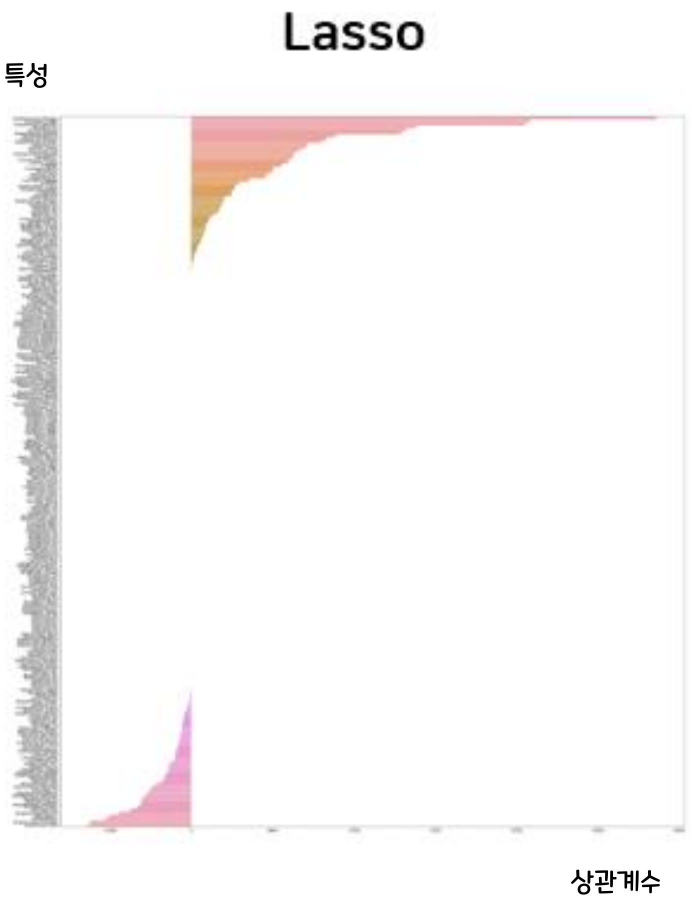


선형 회귀 - 모델 정규화(Model Regularztion)

Kaggle 주택가격 예측 (Ridge vs Lasso)



RMSE = 32152



RMSE = 32592

선형 회귀 - 모델 정규화(Model Regularztion)

구 분	라쏘(Lasso)	릿지(Ridge)	엘라스틱넷(ElasticNet)
적용 규제	L1	L2	L1+L2
특 징	· 중요 하지 않은 변수는 제외 · 특성 간 상관관계가 상대적으로 낮은 경우 사용	· 모든 변수에 같은 비율로 규제를 적용 · 특성 간 상관관계가 상대적으로 높은 경우 사용	· L1규제로 변수를 줄이고 L2규제로 남은 변수들의 영향도를 줄임 · 특성 수가 데이터 수 보다 많을 때 사용

선형 회귀 - 모델 정규화(Model Regularztion)

스마트인재개발원
Smart Human Resources Development

주요 매개변수(Hyperparameter)

scikit-learn의 경우

Ridge(alpha)

Lasso(alpha)

- 규제의 강도 : alpha

alpha값이 커지면 → 규제의 효과가 커짐(과대적합 감소, 오차 증가)

alpha값이 작아지면 → 규제의 효과가 작아짐(과대적합 증가, 오차 감소)
(alpha값이 0이 되면 선형회귀와 같음)

선형 회귀 - 모델 정규화(Model Regularztion)

스마트인재개발원
Smart Human Resources Development

총 105개의 특성을 라쏘 회귀 모델을 만들기 위해 사용

↓

α=1로 설정했더니 105개의 가중치 중에서 101개가 0이 되면서 특성은 단 4개만 사용

↓

훈련셋에서의 점수와 테스트셋에서의 점수를 보니 과소적합

↓

복잡도를 높이기 위해서 α=0.0001로 설정했더니 가중치 중에서 7개만 0이 되면서
94개의 특성이 사용

↓

훈련셋과 테스트셋에서의 점수를 보니 훈련셋은 좋은데 테스트셋은 많이 떨어짐 → 과대적합

↓

다시 복잡도를 낮추기 위해 α=0.1로 설정했더니 105개의 가중치 중에서 72개가 0이 되면서
33개의 특성이 사용 → 훈련셋의 점수와 테스트셋의 점수가 모두 양호

Lasso, Ridge 모델을 이용해 보스턴 집 값 데이터를 활용하여

주택 가격을 예측해 보자

THANK YOU
FOR WATCHING