D&A

ML Session 3차시

회귀 모델

2022 / 09 / 20

D&A 부학회장 김정하

CONTENTS.

01. 선형회귀분석

2022 / 09 / 20
D&A 부학회장 김정하

01. 선형 회귀분석

회귀분석이란?

$$\hat{y} = f(x) \approx y$$

회귀분석은 독립변수 x에 대응하는 종속변수 y와 가장 비슷한 값 $\hat{y}$를 출력하는 함수 $f(x)$를 찾는 과정을 말함

선형회귀분석이란?

$$\hat{y} = \theta_0 + \theta_1 \cdot x_1 + \cdots + \theta_n \cdot x_n$$

벡터식으로 표현

$f(x)$가 선형함수인 회귀모형이면 선형회귀분석이다.

- y_hat : 예측값
- x_i : 회귀모델의 i번째 독립변수
- θ_0 : 편향
- θ_i : i 번째 특성에 대한 (가중치) 파라미터, 단 i>0.

$$\mathbf{x} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

$$\hat{y} = h_\theta(\mathbf{x}) = \theta^T \mathbf{x}$$

회귀분석의 목적

-> 독립변수 x에 대응하는 종속변수 y와 가장 비슷한 값 $\hat{y}$를 출력하는 θ를 찾는다

01. 선형 회귀분석

x y

	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude	target
0	8.3252	41.0	6.984127	1.023810	322.0	2.555556	37.88	-122.23	4.526
1	8.3014	21.0	6.238137	0.971880	2401.0	2.109842	37.86	-122.22	3.585
2	7.2574	52.0	8.288136	1.073446	496.0	2.802260	37.85	-122.24	3.521
3	5.6431	52.0	5.817352	1.073059	558.0	2.547945	37.85	-122.25	3.413
4	3.8462	52.0	6.281853	1.081081	565.0	2.181467	37.85	-122.25	3.422
...	...	...	...	...	...	...	...	...	...
20635	1.5603	25.0	5.045455	1.133333	845.0	2.560606	39.48	-121.09	0.781
20636	2.5568	18.0	6.114035	1.315789	356.0	3.122807	39.49	-121.21	0.771
20637	1.7000	17.0	5.205543	1.120092	1007.0	2.325635	39.43	-121.22	0.923
20638	1.8672	18.0	5.329513	1.171920	741.0	2.123209	39.43	-121.32	0.847
20639	[illegible]	16.0	5.254717	1.162264	1387.0	2.616981	39.37	-121.24	0.894

θ는 각 독립변수 x에 곱해지는 가중치로, 학습을 통해 찾아야하는 값

$$\mathbf{x} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{bmatrix},$$

에 어떤

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

를 곱해야 y(벡터)와 가장 비슷할까?

01. 선형 회귀분석

x y

	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude	target
0	8.3252	41.0	6.984127	1.023810	322.0	2.555556	37.88	-122.23	4.526
1	8.3014	21.0	6.238137	0.971880	2401.0	2.109842	37.86	-122.22	3.585
2	7.2574	52.0	8.288136	1.073446	496.0	2.802260	37.85	-122.24	3.521
3	5.6431	52.0	5.817352	1.073059	558.0	2.547945	37.85	-122.25	3.413
4	3.8462	52.0	6.281853	1.081081	565.0	2.181467	37.85	-122.25	3.422

-> 1. y가 있는 데이터로 y를 가장 잘 설명하는 식 $\hat{y} = \theta_0 + \theta_1 \cdot x_1 + \cdots + \theta_n \cdot x_n$ 를 완성시켜 최적의 θ를 찾는다.

20636	2.5568	18.0	6.114035	1.315789	356.0	3.122807	39.49	-121.21
20637	1.7000	17.0	5.205543	1.120092	1007.0	2.325635	39.43	-121.22
20638	1.8672	18.0	5.329513	1.171920	741.0	2.123209	39.43	-121.32
20639	2.3886	16.0	5.254717	1.162264	1387.0	2.616981	39.37	-121.24

-> 2. 1에서 찾은 회귀식에 x값만 있는 데이터를 적용시켜 y를 예측한다.

$$\hat{y} = \theta_0 + \theta_1 \cdot x_1 + \cdots + \theta_n \cdot x_n$$

-> y를 잘 나타내는 θ를 찾아 다른 x값을 넣어도 y를 잘 예측할 수 있게 만든다!

01. 선형 회귀분석 - 경사하강법

Loss function

손실함수라고 하며, 실제값과 예측값의 차이를 특정 함수로 나타내어
이 함수를 최소화 시키는 방향으로 모델의 학습이 진행됨
회귀모델에서는 MSE를 Loss function으로 주로 사용

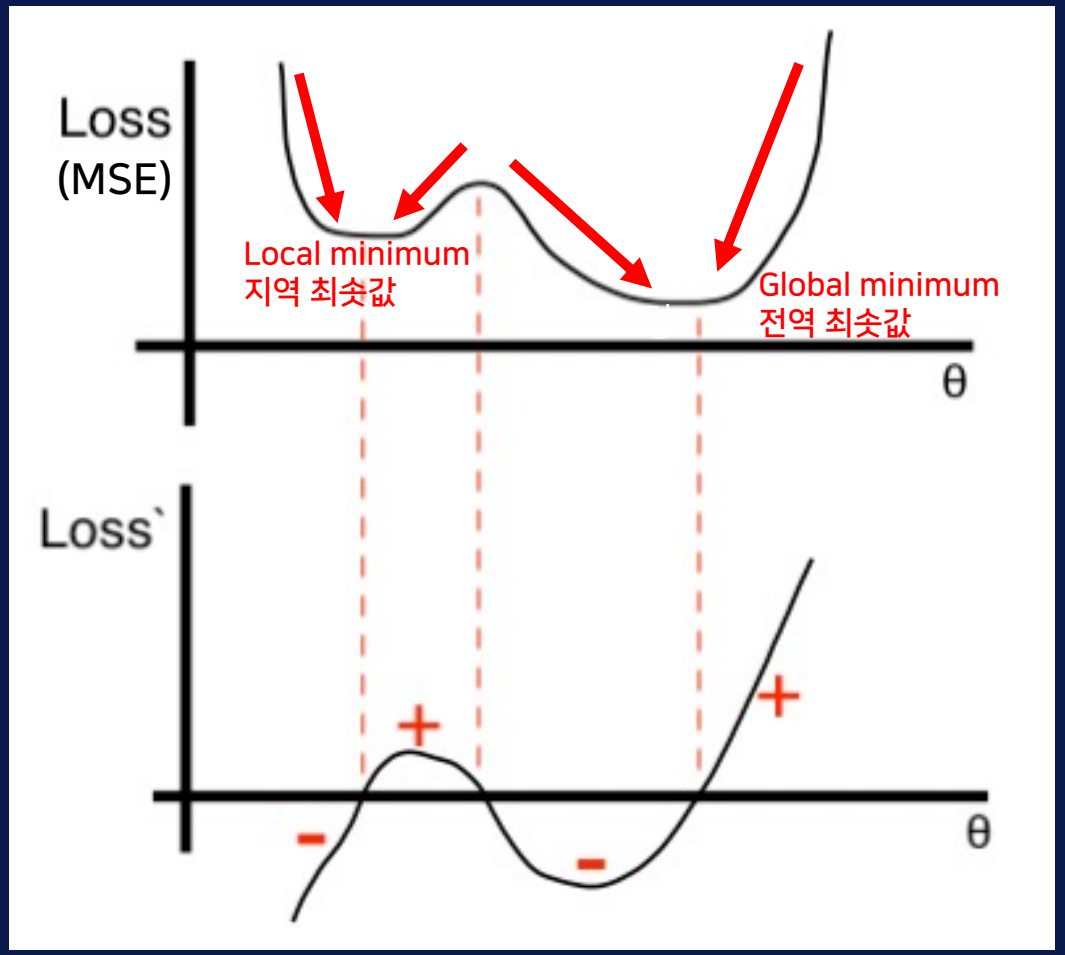
MSE

평균 제곱 오차로, 각 실제값과 예측값의 오차를 제곱한 값들을 평균한 값

$$\mathrm{MSE}(\theta) := \mathrm{MSE}(\mathbf{X}, h_\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(\theta^T \mathbf{x}^{(i)} - y^{(i)}\right)^2$$

01. 선형 회귀분석 - 경사하강법

이론적인 경사하강법



-> Loss의 미분값이 음수일 때는 θ가 커지게,
양수일 때는 θ가 작아지게 학습!

기울기 반대 방향으로 η 배 만큼 이동

$$\theta_{i,j+1} = \theta_{i,j} - \eta \frac{\partial}{\partial \theta_{i,j}} J(\theta)$$

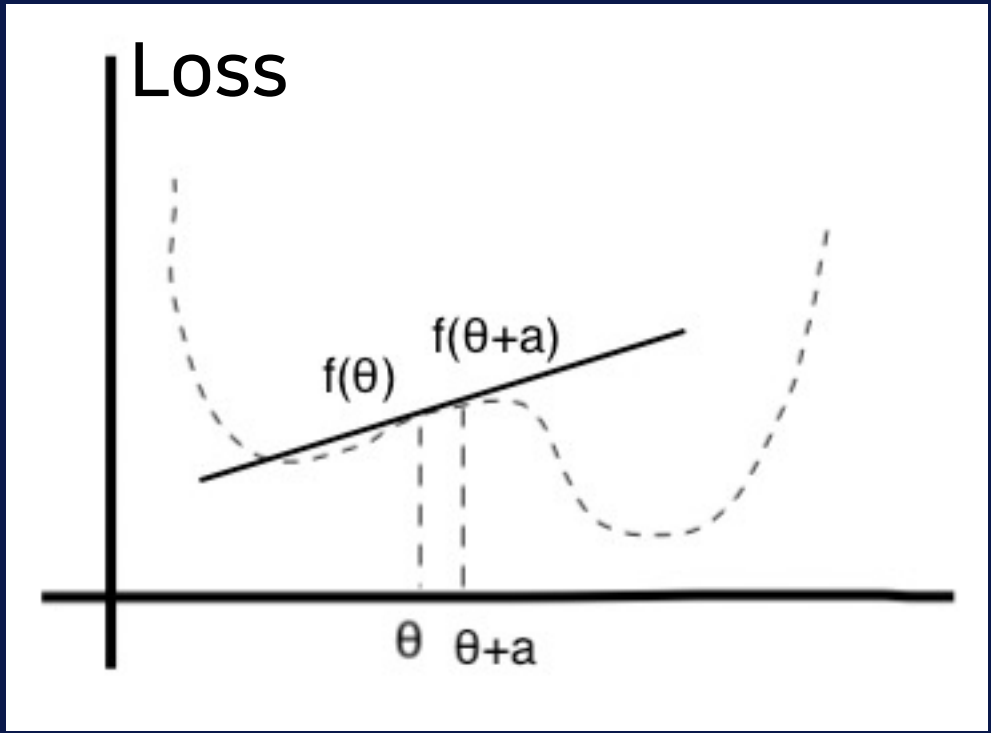
학습률

- $\theta_{i,j+1}$: 새롭게 갱신된 가중치
- $\theta_{i,j}$: 갱신 전 가중치
- $\frac{\partial}{\partial \theta_{i,j}} J(\theta)$: 해당 지점에서의 기울기

-> θ를 업데이트할 때, Loss의 미분값을 활용

01. 선형 회귀분석 - 경사하강법

실제 모델 속 경사하강법



-> θ를 대입해봐야 Loss를 구할 수 있기 때문에, 그래프 상에서 한 점씩 알아가며 수치미분을 통해 경사하강법 진행

수치미분

$$\frac{df(\theta)}{d\theta} = \lim_{a \to 0} \frac{f(\theta + a) - f(\theta)}{a}$$

- a는 아주 작은 값으로 , 10e-4가 좋다고 알려져있다.

중앙차분

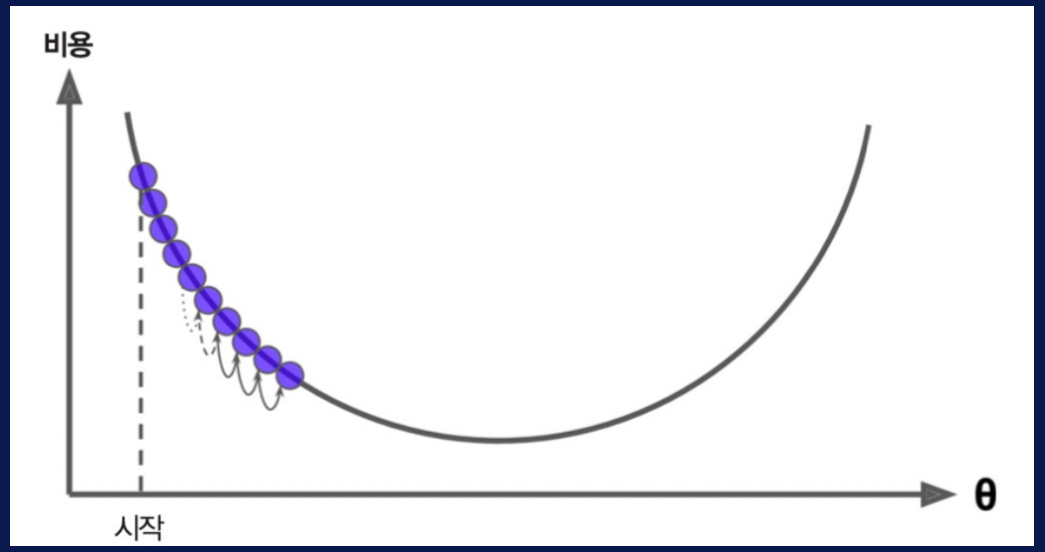
$$\frac{df(\theta)}{d\theta} = \lim_{a \to 0} \frac{f(\theta + a) - f(\theta - a)}{2a}$$

- 실제 기울기와 수치미분값의 오차를 줄이기 위해 중앙차분 사용

01. 선형 회귀분석 - 경사하강법

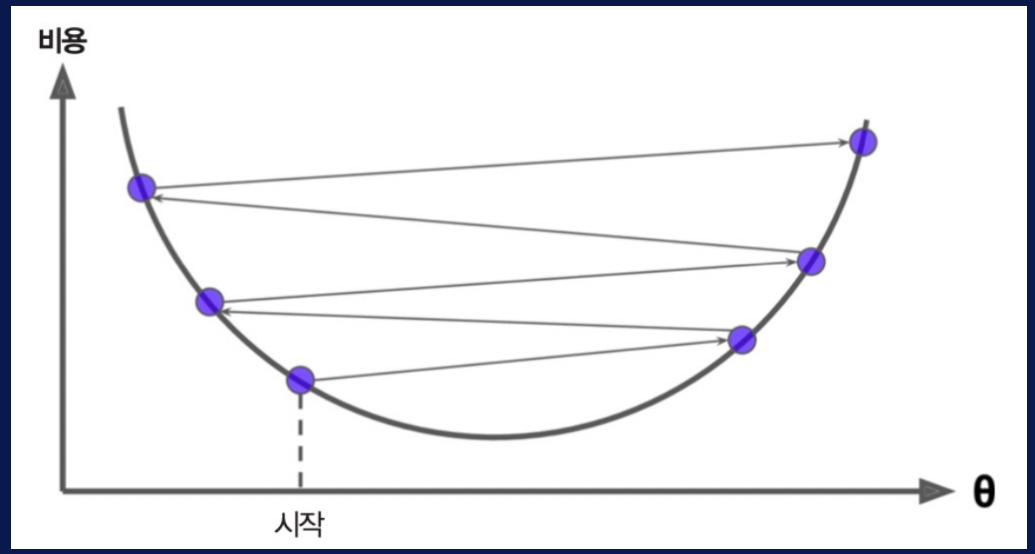
학습률 - θ를 얼마나 업데이트 시킬지 정하는 하이퍼파라미터

학습률이 너무 작을 때



-> 최솟값에 너무 느리게 수렴함

학습률이 너무 클 때



-> 손실함수가 수렴하지 않음

01. 선형 회귀분석 – 경사하강법_SGD

SGD란?

확률적 경사하강법(Stochastic gradient descent)으로, 회귀 모델이 가중치를 한 번 업데이트할 때 사용하는 데이터를 무작위로 한 개의 샘플을 선택해 경사하강법을 진행함

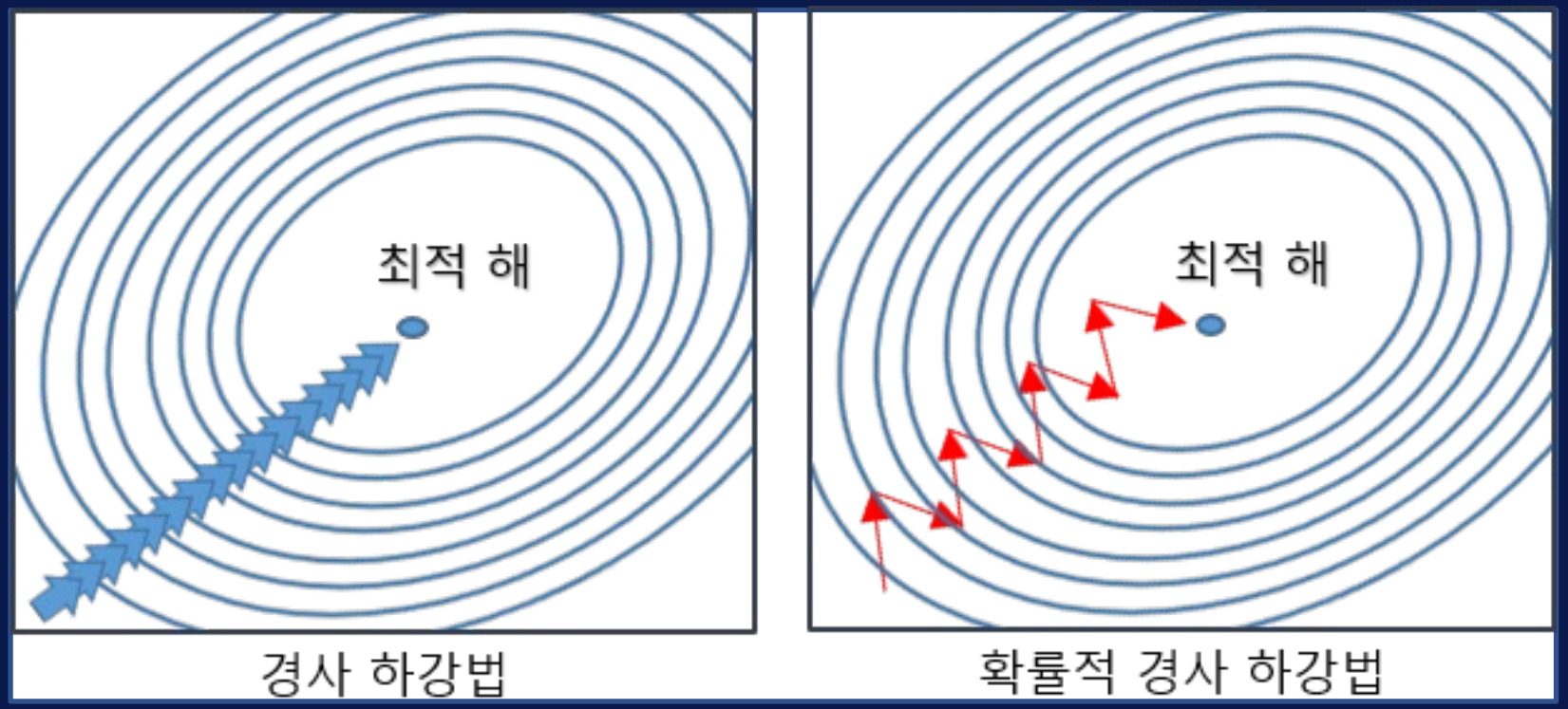
-> 매 반복에서 하나의 샘플을 사용해 속도가 빠름

-> 확률적(무작위)로 샘플을 선택해 알고리즘을 수행하기 때문에 일반 경사하강법보다 불안정함

-> Loss function이 불규칙한 특성을 띄고 있을 때 지역 최솟값(local minimum)을 뛰어넘을 수 있기 때문에 전역 최솟값(global minimum)을 찾을 가능성이 큼

01. 선형 회귀분석 – 경사하강법_SGD

경사하강법의 학습 진행 과정



-> SGD는 샘플마다의 편차로 불규칙적으로 움직이며 Loss 최솟값을 찾는 중

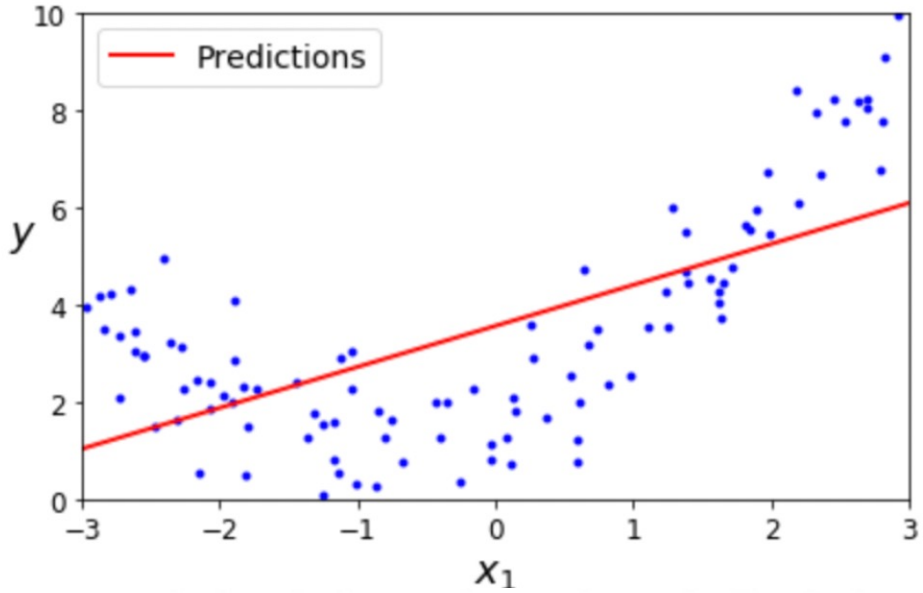
02. 다항회귀

2022 / 09 / 20
D&A 부학회장 김정하

02. 다항 회귀

비선형적인 데이터셋에 선형 회귀 모델 적용 결과

$$\hat{y} = \theta_0 + \theta_1 x_1$$

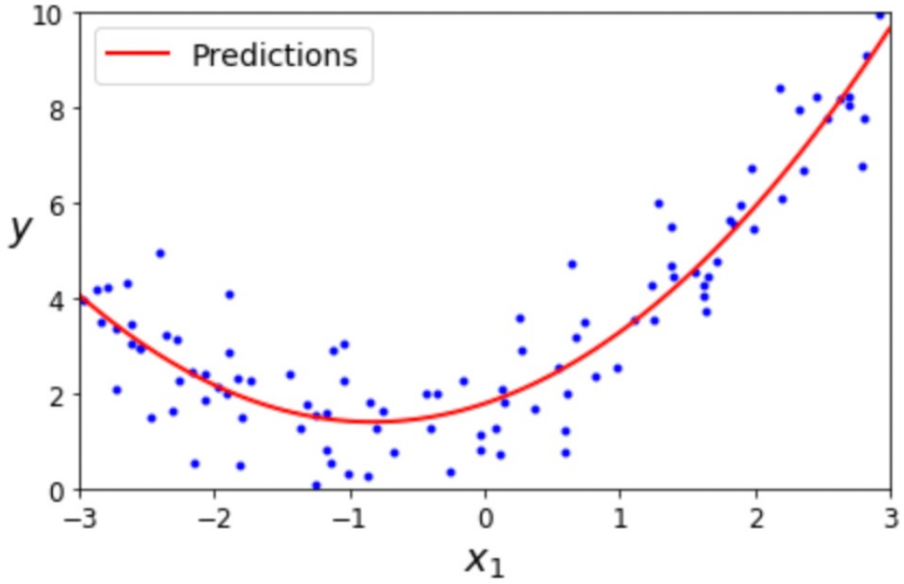


선형 회귀 모델로 학습시킨 결과

-> 독립변수의 일차항만을 가진 선형함수는 곡선의 비선형적인 실제값들을 잘 나타낼 수 없음

비선형적인 데이터셋에 2차 다항식 모델 적용 결과

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$



x_1^2 에 해당하는 특성 x_2 를 새로이 추가한 후에 선형 회귀 모델을 학습시킨 결과

-> 변수의 차수가 늘어남에 따라 실제값과의 오차가 줄어듬

02. 다항 회귀

Scikit-learn의 PolynomialFeatures 모듈

```
PolynomialFeatures(degree=d, include_bias=False)
```

-> 다항식에 포함되어야하는 특성(독립변수)들을 생성해주는 변환기

-> degree는 다항식의 차수를 뜻한다.

Example) x1과 x2 두 개의 독립변수에 대해 degree = 3으로 다항회귀식을 생성하면?

-> $(x_1 + x_2)^2$ 과 $(x_1 + x_2)^2$ 의 항들을 새로운 특성으로 추가하여

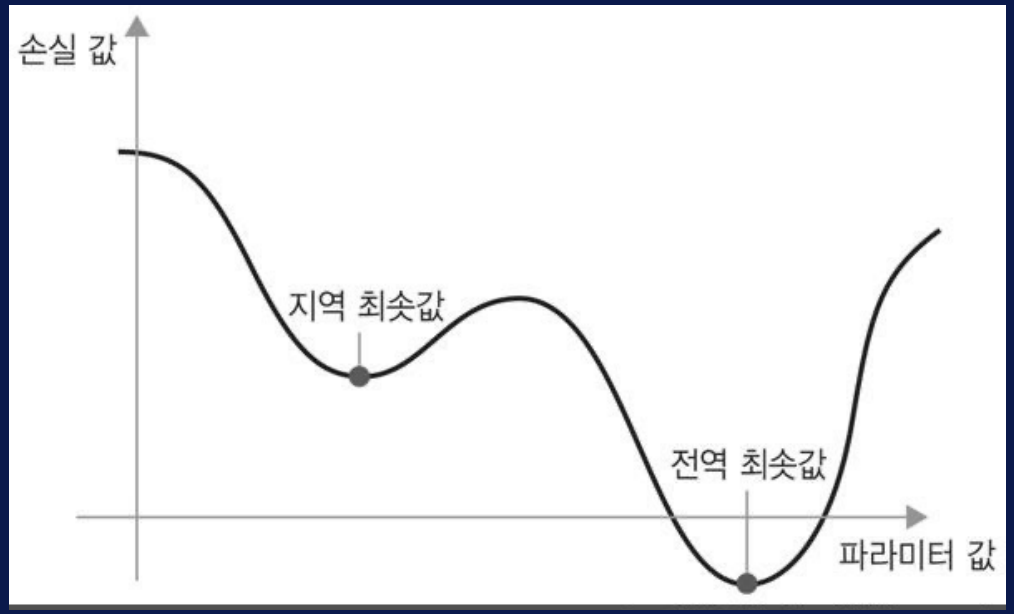
$$x_1^2,\ x_1x_2,\ x_2^2,\ x_1^3,\ x_1^2x_2,\ x_1x_2^2,\ x_2^3$$

-> 이 같은 항들을 추가!

03. 규제가 있는 선형회귀

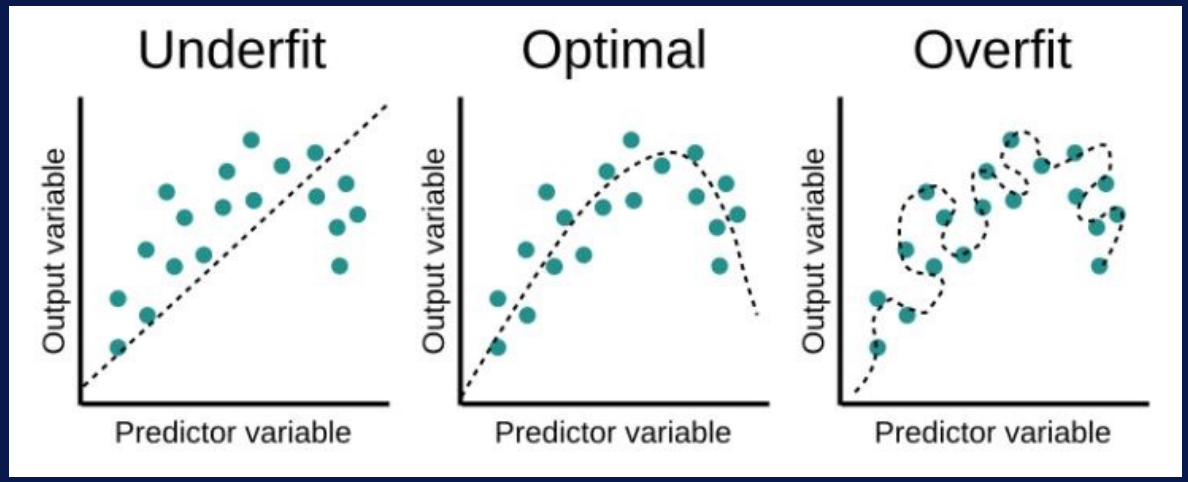
2022 / 09 / 20
D&A 부학회장 김정하

03. 회귀모델의 규제



전역 최솟값을 찾더라도, 이는 학습데이터에 맞춰진 전역 최솟값임..!

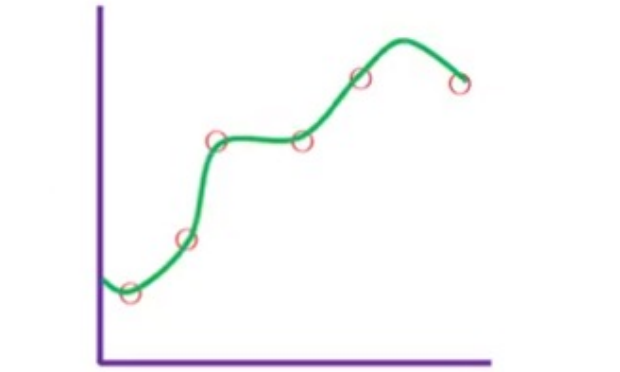
Overfitting



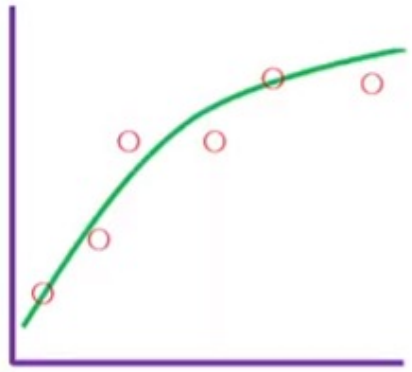
3번 그래프의 Loss의 값이 제일 적지만,
학습데이터에 과도하게 학습되어 검증데이터나 실제 적용할 데이터에 맞지 않을 가능성이 큼

03. 회귀모델의 규제

규제를 통한 과적합 줄이기



$$\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$$



$$\beta_0 + \beta_1 x + \beta_2 x^2$$

-> 모든 독립변수들의 영향력이 존재(가중치가 0 초과)할 때 overfitting

-> 독립변수의 개수를 줄이거나(해당 가중치가 0) 영향력을 줄이면 overfitting감소

03. 회귀모델의 규제

규제를 통한 과적합 줄이기

기존의 손실함수 MSE

$$\mathrm{MSE}(\theta) := \mathrm{MSE}(\mathbf{X}, h_\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(\theta^T \mathbf{x}^{(i)} - y^{(i)}\right)^2$$

목적 => 이 손실함수를 최소화 시키자!

$$min(MSE(\theta) + penalty)$$

최소화시키고자하는 손실함수에 규제(penalty)를 추가해 독립변수에 규제를 주어 Overfitting을 감소시키자!

03. Ridge 회귀모형

Ridge 회귀모형의 손실함수

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

θ에 대한 L2규제 추가

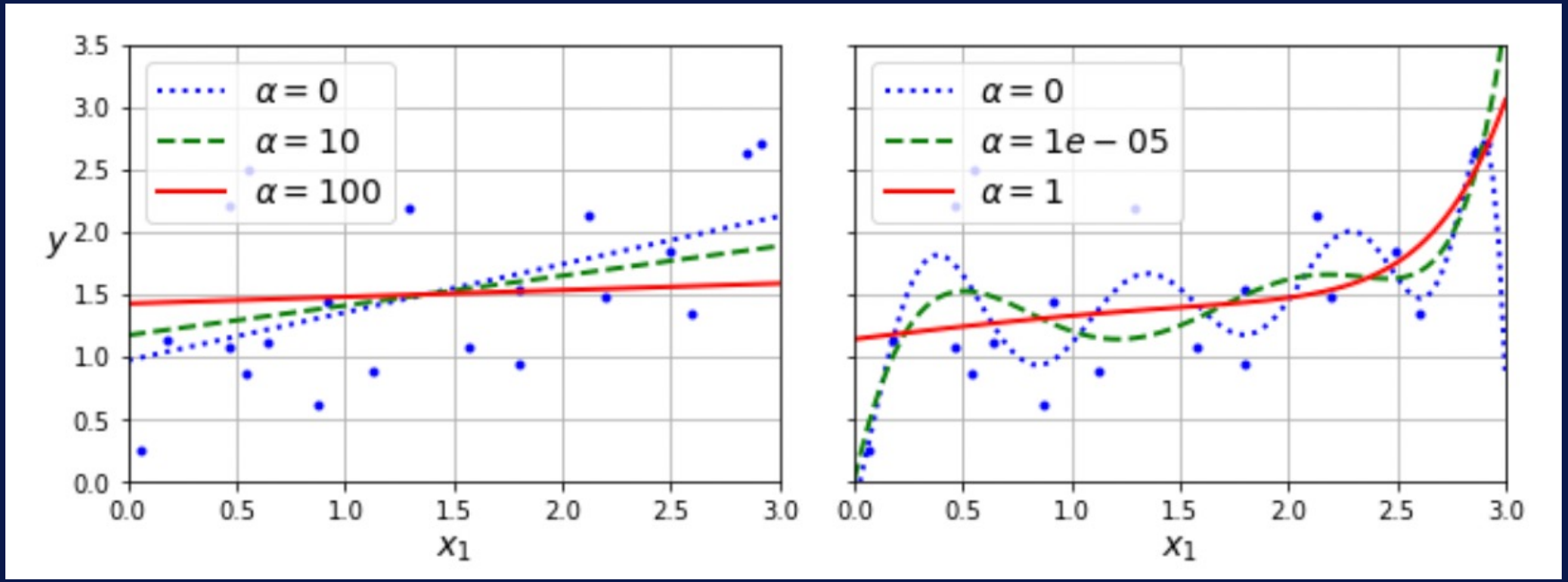
θ의 제곱합에 대한 규제를 추가하여
θ의 전체적인 크기에 제약을 줌

기존의 손실함수 MSE

규제의 정도를 조절하는 하이퍼 파라미터
α가 0이면? 기존의 선형회귀모델의 손실함수

03. Ridge 회귀모형

α 는 규제의 정도를 조절하는 하이퍼 파라미터



α 가 커질수록 회귀계수에 대한 규제가 커짐
-> 더 일반화된(overfitting감소) 회귀식으로 변화

03. Lasso 회귀모형

θ에 대한 L1규제 추가

θ의 절대값합에 대한 규제를 추가하여
θ의 전체적인 크기에 제약을 줌

Lasso 회귀모형의 손실함수

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

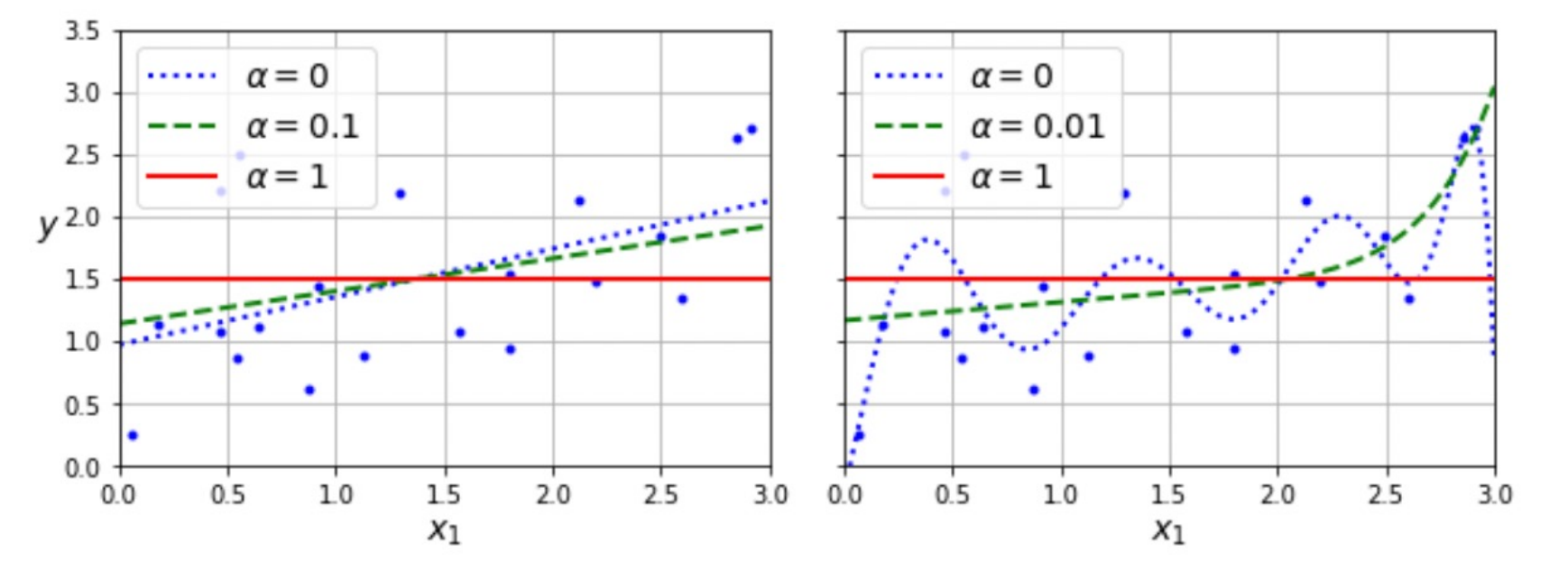
기존의 손실함수 MSE

$|\theta_i|$ 의 미분값이 1 또는 -1이 되어

규제의 정도를 조절하는 하이퍼 파라미터
α 가 0이면? 기존의 선형회귀모델의 손실함수

03. Lasso 회귀모형

α 는 규제의 정도를 조절하는 하이퍼 파라미터



α 가 커질수록 회귀계수에 대한 규제가 커짐
-> 더 일반화된(overfitting감소) 회귀식으로 변화
-> 중요하지 않은 특성에 대해 θ가 0에 빠르게 수렴함

03. Elastic Net

a는 규제의 정도를 조절하는 하이퍼 파라미터

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

Lasso회귀와 Ridge회귀의 규제항을 동시에 사용
하이퍼파라미터 r을 사용해 두 규제항의 혼합비율을 조절

기존의 손실함수 MSE

θ에 대한 L1규제(Lasso)

θ에 대한 L2규제(Ridge)

03. 규제가 있는 회귀모델

언제 어떤 규제 사용 모델을 쓸까?

- 일반적으로는 회귀 모델에 규제를 사용할 때는 Ridge가 추천됨

- 유용한 특성이 그렇게 많지 않다고 판단되면, Lasso 또는 ElasticNet 사용
  -> Lasso의 규제항이 유용하지 않은 특성을 없애 주기 때문

-

04. 로지스틱 회귀

2022 / 09 / 20
D&A 부학회장 김정하

04. 로지스틱 회귀

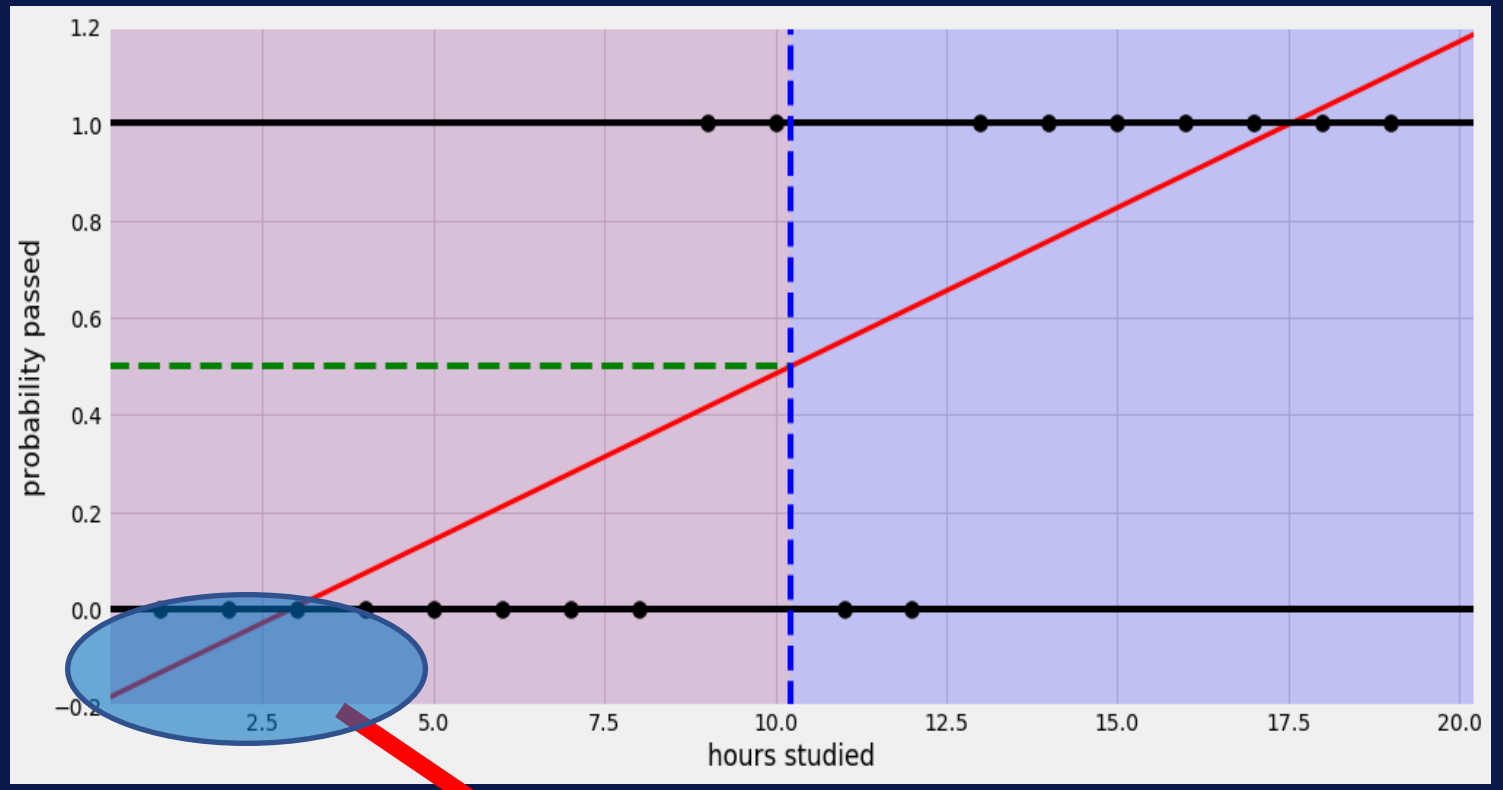
로지스틱회귀 (Logistic Regression)

회귀를 사용하여 데이터가 어떤 범주에 속할 확률을 0에서 1 사이의 값으로 예측하고 그 확률에 따라 가능성이 더 높은 범주에 속하는 것으로 분류해주는 지도 학습 알고리즘

ex) 스팸 메일일 확률이 0.5 이상이면 스팸메일로, 그 미만이면 일반 메일로 분류

04. 로지스틱 회귀

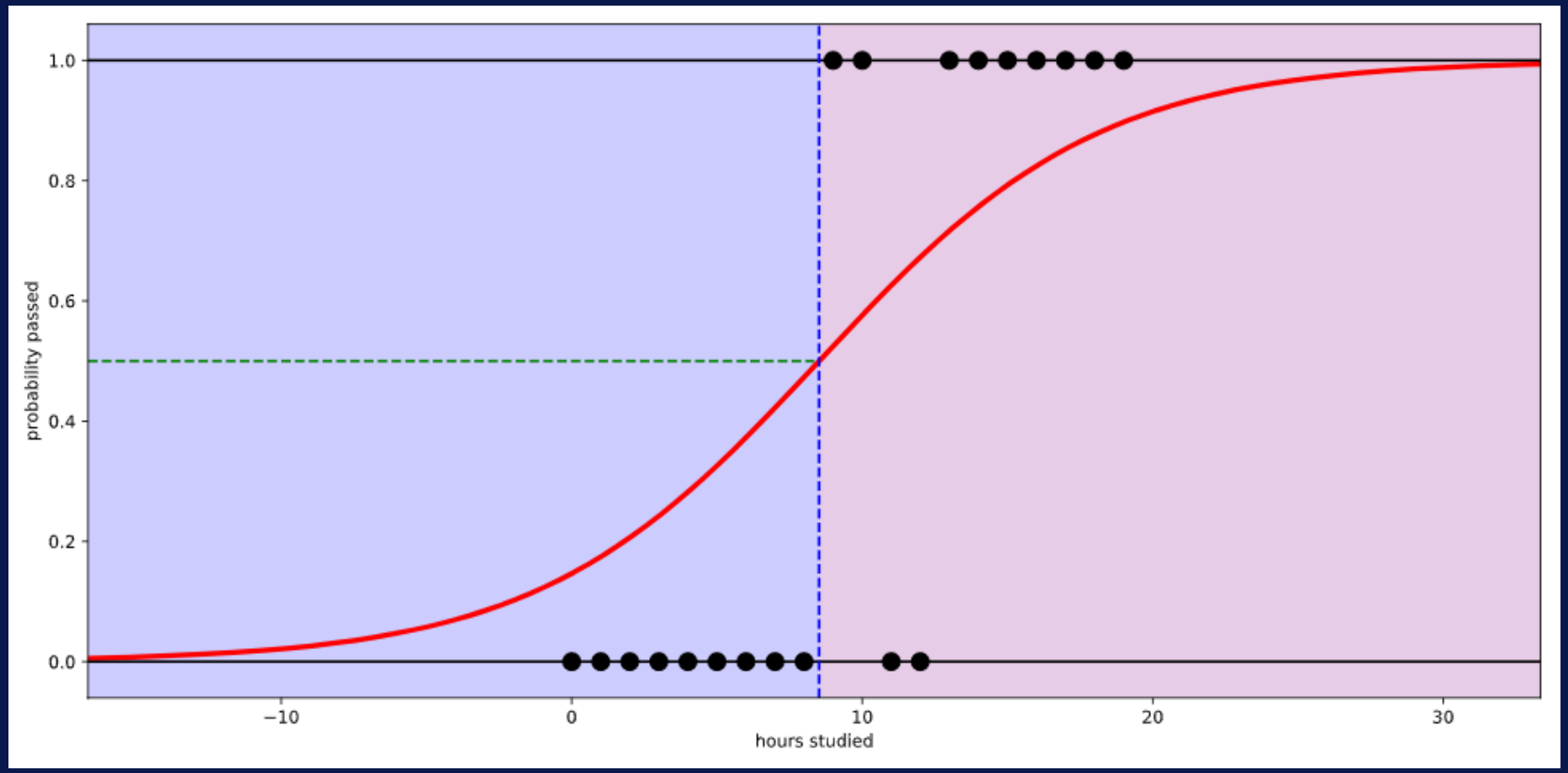
선형회귀로 확률값을 예측한다면?



공부시간이 2시간 미만이면 합격확률이 음수가 됨

04. 로지스틱 회귀

로지스틱 회귀로 확률값을 예측한다면?

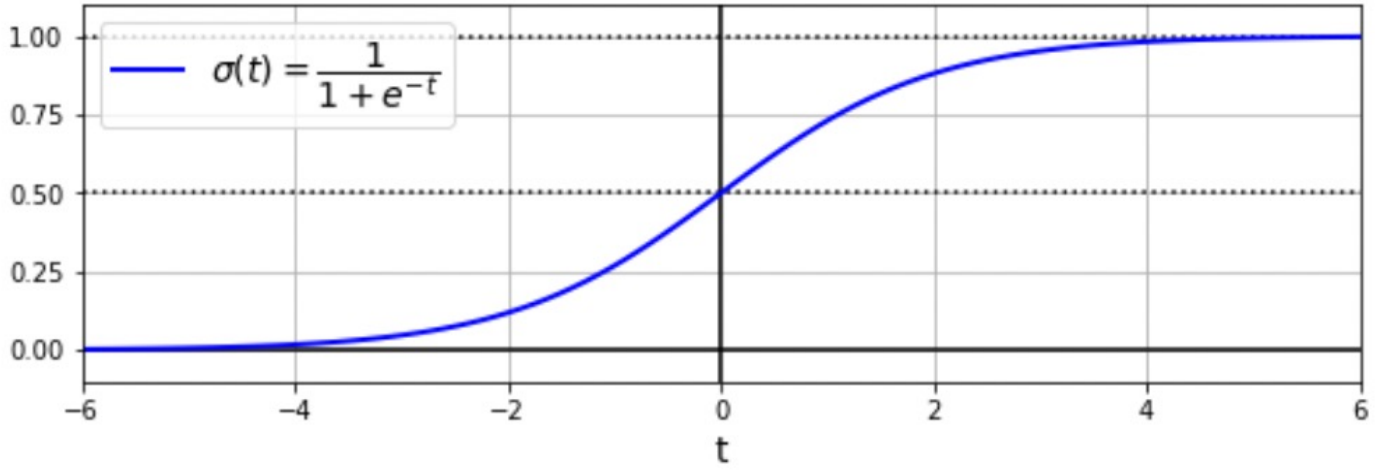


시험에 합격할 확률이 0과 1사이의 값으로 정해짐

04. 로지스틱 회귀

sigmoid 함수

$$\sigma(t) = \frac{1}{1 + \exp(-t)}$$



회귀 모델의 최종값을 sigmoid함수에 넣어 0~1의 확률값으로 만듬

04. 로지스틱 회귀

결과값 분류

$$\hat{p} = h_\theta(\mathbf{x}) = \sigma(\theta^T \mathbf{x}) = \sigma(\theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n)$$

일 때,

$$\hat{y} = \begin{cases} 0 & \text{if } \hat{p} < 0.5 \\ 1 & \text{if } \hat{p} \geq 0.5 \end{cases}$$

0.5를 기준으로 두 개의 클래스로 분류

04. 로지스틱 회귀

로지스틱 회귀의 Loss function – log loss

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\ \log(\hat{p}^{(i)}) + (1-y^{(i)})\ \log(1-\hat{p}^{(i)})\right]$$

실제 분류가 1일 때 첫 번째 항만!
-> log(p)가 커지도록, 즉 p가 커지게 학습됨

실제 분류가 0일 때 두 번째 항만!
-> log(1-p)가 커지도록, 즉 p가 작아지게 학습됨

04. 로지스틱 회귀

로지스틱 회귀 정리

1. 선형회귀의 결과값을 시그모이드 함수에 적용해 확률값(0~1)로 계산

$$\hat{p} = h_\theta(\mathbf{x}) = \sigma(\theta^T \mathbf{x}) = \sigma(\theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n)$$

확률값 시그모이드 함수 선형회귀 결과값

2. Log Loss (로그 손실)을 Loss function으로 사용,
   예측 확률값을 실제값과 비슷해지도록 학습

   -> 경사하강법을 위한 손실함수의 미분 과정은 직접 해보기^^(재미있음)

첨부자료 출처

폰트

네이버 글꼴 모음 _ 나눔 스퀘어 사용
출처 : https://hangeul.naver.com/font

D&A

ML session 3차시 회귀모델

Thank You.

2022 / 09 / 20

D&A 부학회장 김정하