Logistic Regression
분류

머신러닝/딥러닝

임 경 태

Week	Chapter	Contents
1	1, 2장	강의 소개, 파이썬 복습
2	1, 3장	파이썬 복습, Numpy, Pandas
3	1, 4장	딥러닝을 위한 미분
4	5장	회귀
5	5장	분류
6	6장	XOR문제
7	7장	딥러닝
8	1~7장	중간고사
9	8장	MNIST 필기체 구현 (팀 프로젝트)
10	9장	오차역전파
11	11장	합성곱 신경망(CNN)
12	12장	순환 신경망(RNN)
13	10장	자율주행 (Collision Avoidanve, Transfer Learning)
14	11장	자율주행 (Load Following)
15	8~12장	기말고사 (or 프로젝트 발표)

CONTENTS

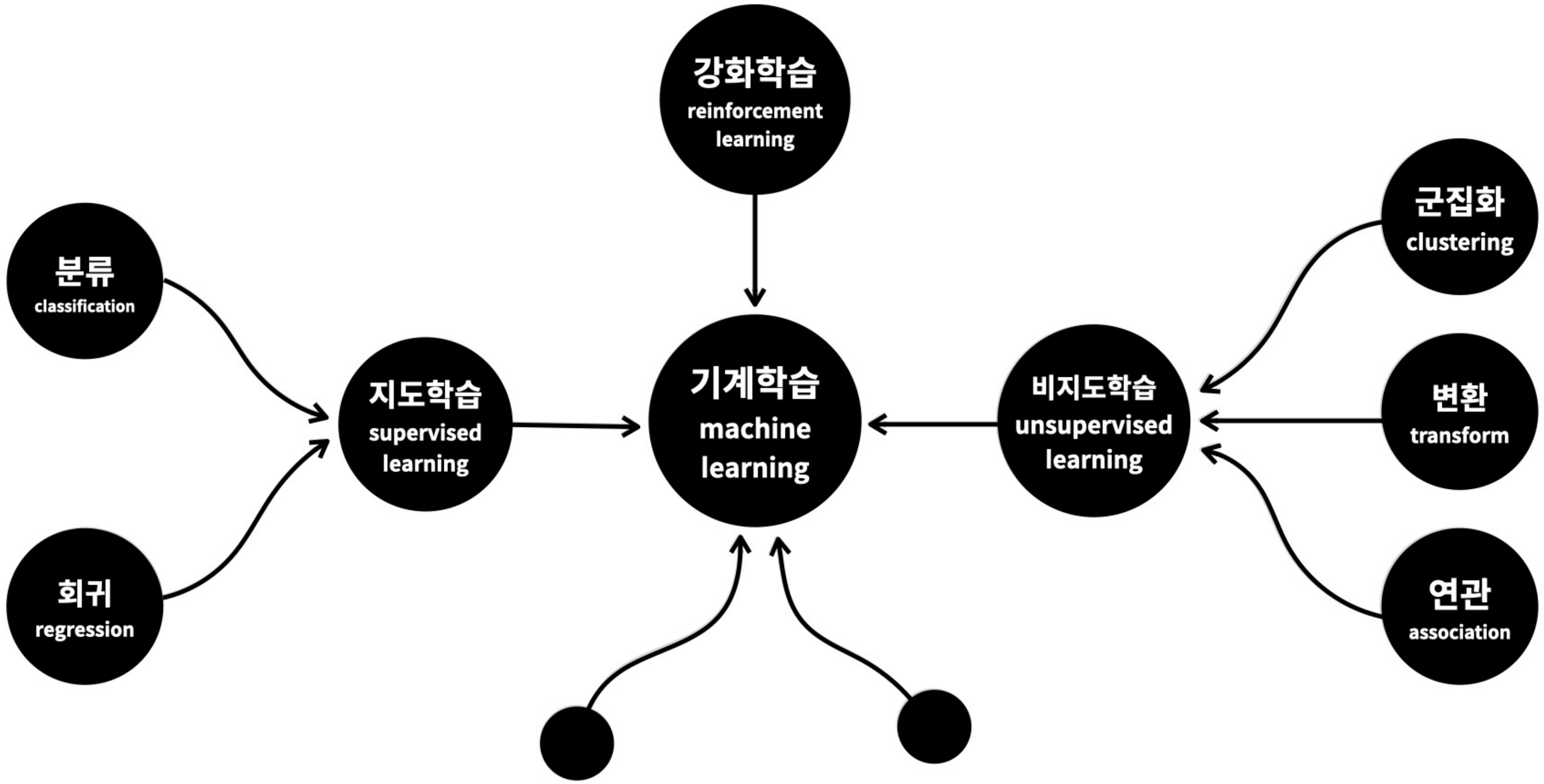
목적 : 지도학습의 대표적인 알고리즘인 분류에 대한 학습

목표 : Classification, Sigmoid, Cross-entropy에 대한 이해

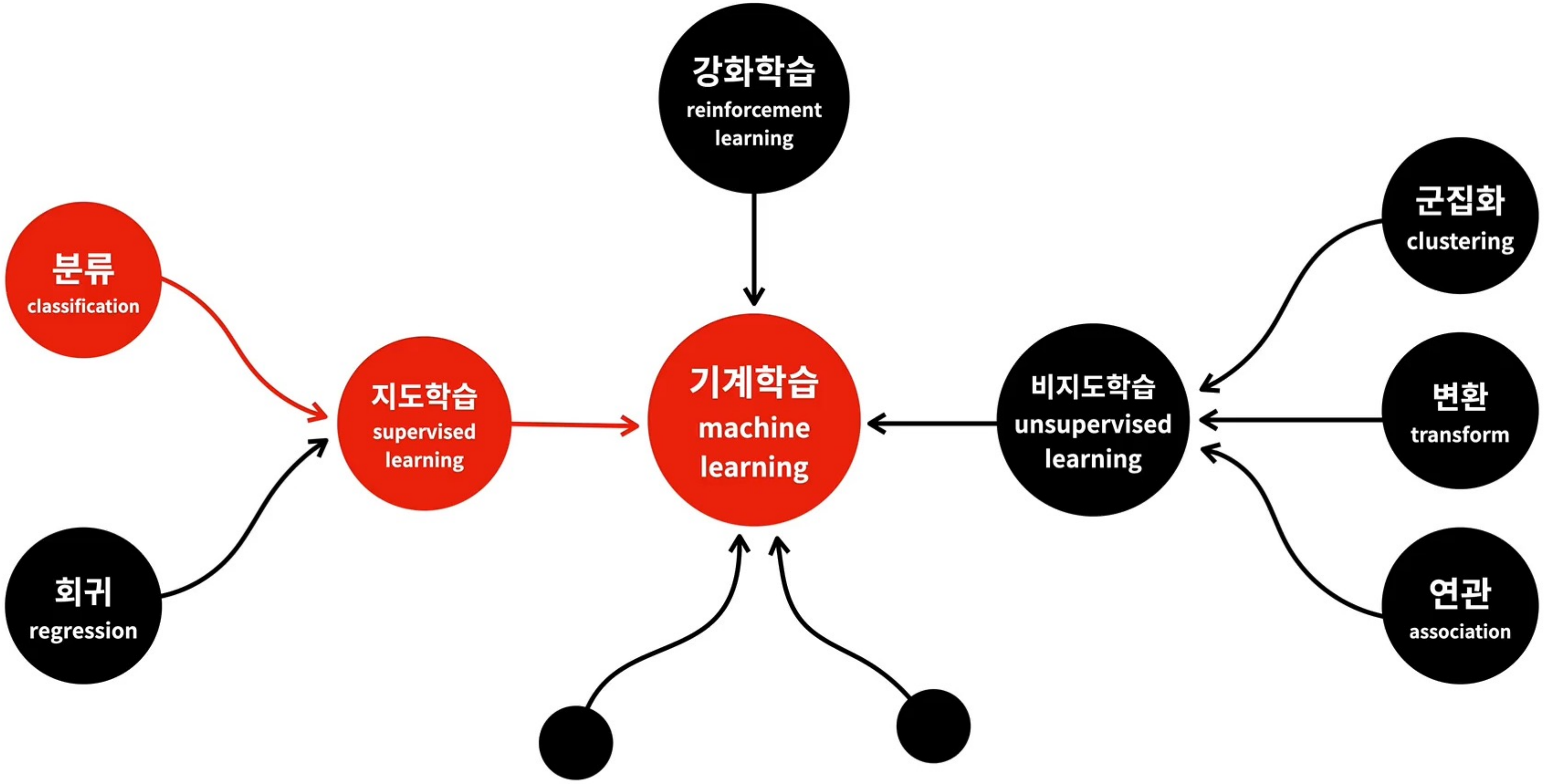
내용 : 분류 개념, Logistic Regression, Sigmoid, Cross-entropy

CONTENTS

머신러닝과 분류


강화학습
reinforcement learning
분류
classification
지도학습
supervised learning
기계학습
machine learning
비지도학습
unsupervised learning
군집화
clustering
변환
transform
회귀
regression
연관
association

머신러닝과 분류



강화학습
reinforcement learning
분류
classification
지도학습
supervised learning
기계학습
machine learning
비지도학습
unsupervised learning
군집화
clustering
변환
transform
연관
association
회귀
regression

분류 개념과 특징

- 정답 레이블이 있는 데이터로 지도학습을 한 후 새로 입력된 데이터가 어느 그룹에 속하는지 알아내는 것.
- 회귀는 연속적인 값을, 분류는 이산적인 값(범주형)을 예측하는 것에서 차이가 있음
- 로지스틱 회귀, KNN, SVM, Naïve Bayes, 의사결정 트리 등의 알고리즘이 대표적임

공부 시간	합/불
9	불합격
14	불합격
21	불합격
27	합격
32	합격
37	합격

공부 시간 → 시험 합격/불합격

① input

② learning

Classification

③ ask (공부 시간)

공부 시간
55
60
78

④ predict

Train data에 대한 시험 성적 예측

공부 시간	합/불
55	??
60	??
78	??

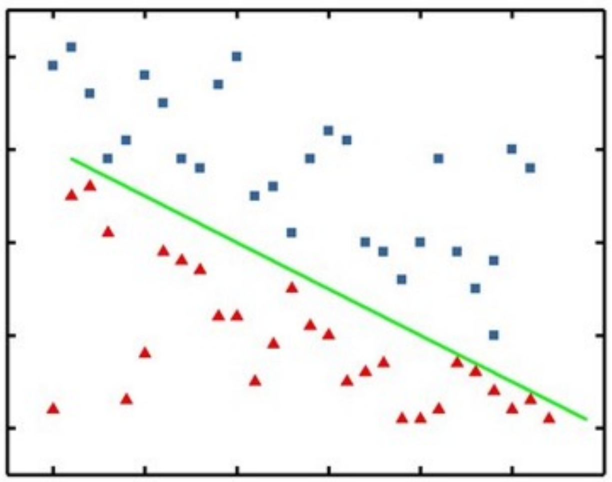
분류 문제의 구분

- Binary Classification (이진 분류) 문제는 무엇인가?
  - 스팸메일 검출: Yes, No
  - 영화 추천: 추천, 비추천
  - 수능 합/불: 합격, 불합격
  - Where y={0, 1}

- Multinominal Classification (다중 분류) 문제?
  - 수능 등급 분류: 1등급, 2등급, 3등급, 4등급
  - 객체 인식: 고양이, 강아지, 사자
  - 감정 인식: 기쁨, 슬픔, 냉소, 분노
  - Where y={0, 1, 2, .. n}

CONTENTS

Logistic Regression의 분류 방법



(a) Logistic Regression

1. Training Data 의 특성과 분포를 나타내는 최적의 직선을 찾는다.
2. 직선을 기준으로 분류 (위 = 1, 아래 = 0 등)
3. 미지의 데이터(Test Data)를 분류

Logistic Regression은 정확도가 높은 Classification 알고리즘

공부 시간	합/불
9	불합격
37	합격

학습 데이터

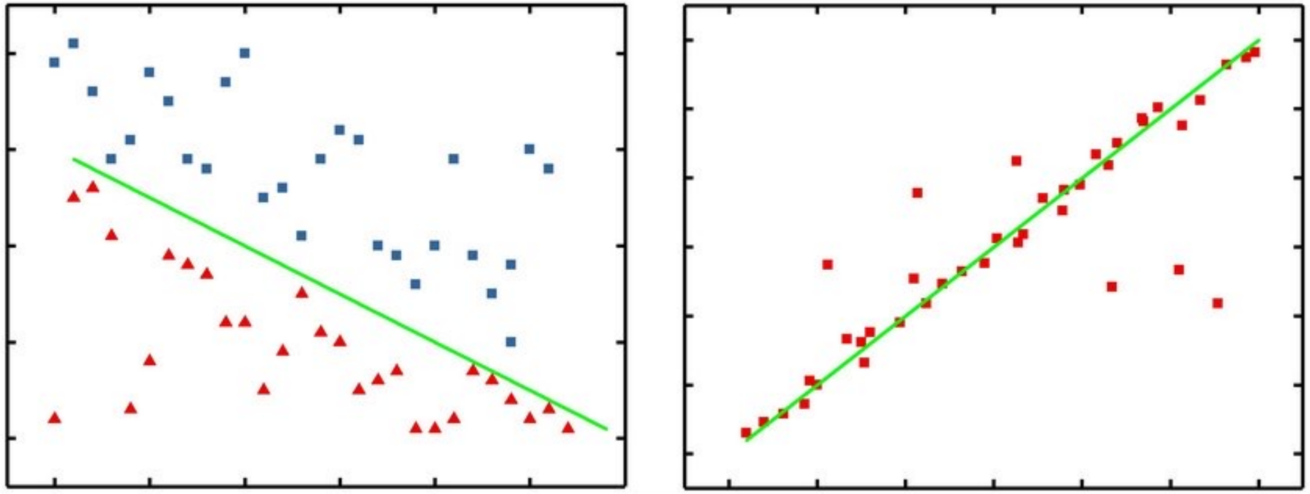
→ Regression → Classification →

공부 시간	합/불
9	??
37	??

예측

Logistic Regression VS Linear Regression

Linear Regression은 연속된 값을 예측하는 반면,
Logistic Regression은 discrete한 값을 예측한다.
→ classification



(a) Logistic Regression (b) Linear Regression

Logistic Regression (개요)

❑ 점수(정수)가 아닌 Binary(Pass/fail) 분류 문제

- 우리반 학생들의 H 대학 지원 결과가 아래와 같다. 나는 7,3 했는데 합격할 수 있을까?

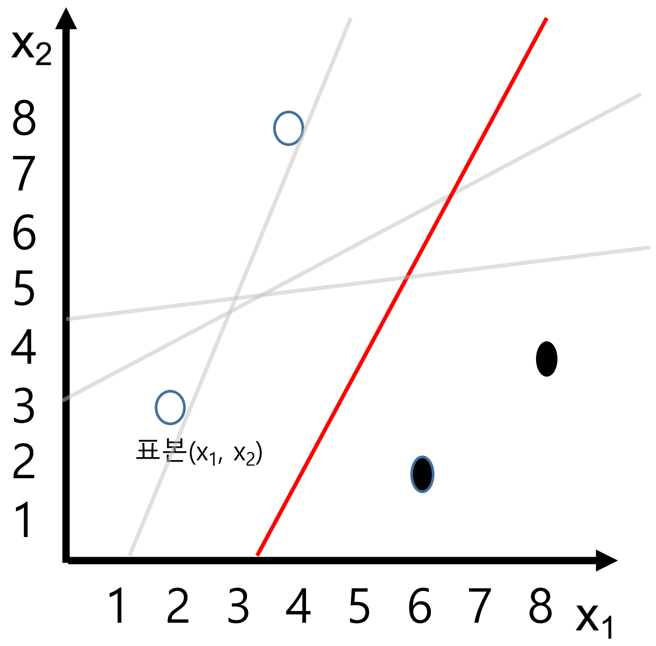
이름	수학 공부시간	영어 공부시간	합/불
야쓰오	2	3	불합격
오공	4	8	불합격
블리츠	6	2	합격
페이커	8	4	합격

Logistic Regression (개요)

❑ 합격 불합격을 어떻게 나눌 수 있을까?

- **로지스틱 회귀**의 목적은 일반적인 회귀 분석의 목표와 동일하게 종속 변수와 독립 변수간의 관계를 구체적인 함수로 나타내어 향후 예측 모델에 사용하는 것이다 (위키)

이름	수학 (x_1) 공부시간	영어 (x_2) 공부시간	합/불 (y)
야쓰오	2	3	불합격
오공	4	8	불합격
블리츠	6	2	합격
페이커	8	4	합격

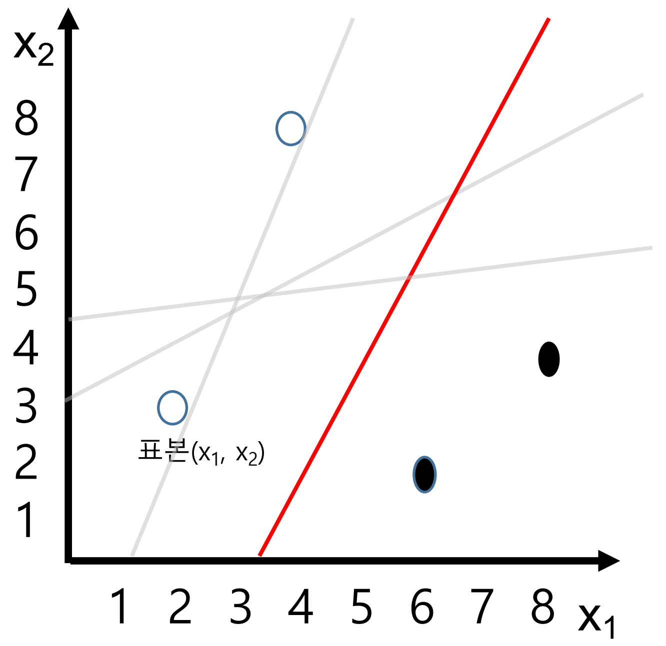


Logistic Regression (개요)

❑ 예측 값 합격/불합격은 숫자가 아닌데?

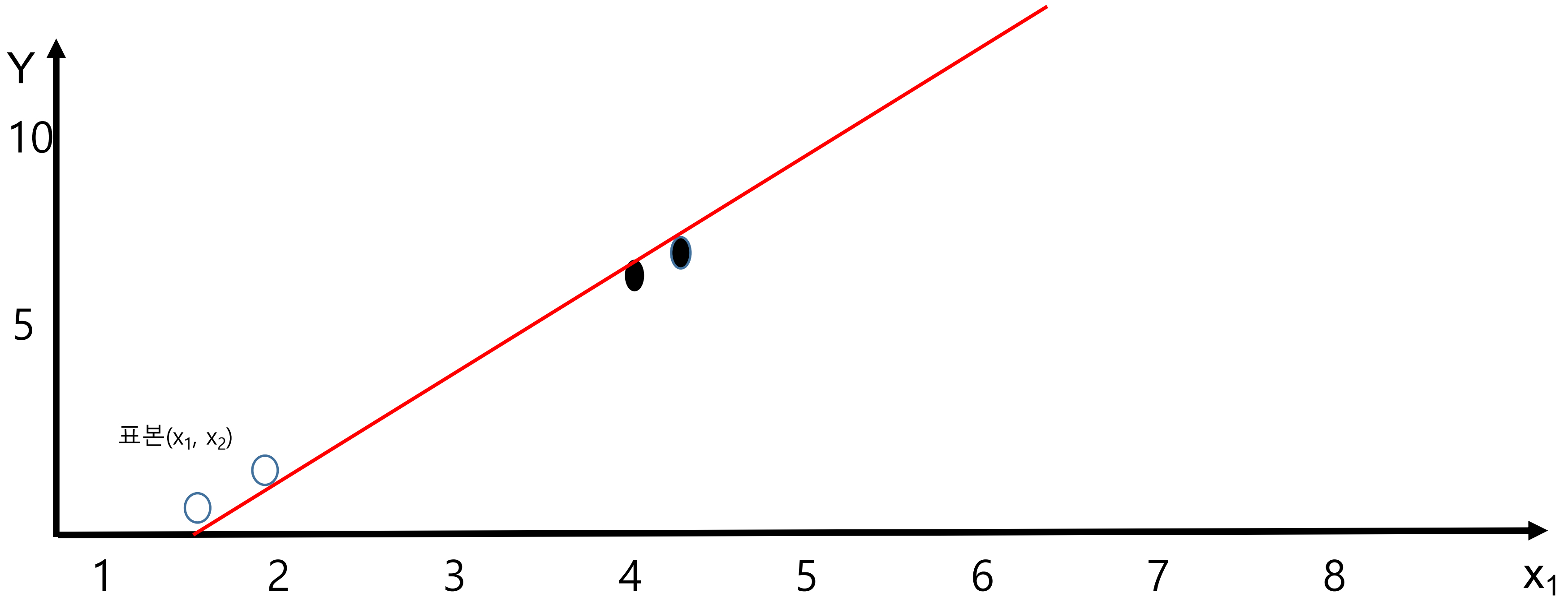
- ❑ 숫자로 만들지 뭐! 종속 변수 y 결과의 범위를[0,1]로 제한해서 0이면 불합격 1이면 합격으로 만들자!
- ❑ 종속 변수 y가 이진적이기 때문에 조건부 확률(P(y│x))의 분포가 이항 분포를 따른다
- ❑ 종속 변수 y의 결과는 0과 1, 두 개의 경우만 존재하는데 반해, 단순 선형 회귀를 적용하면 범위[0,1]를 벗어나는 결과가 나오기 때문에 오히려 예측의 정확도만 떨어뜨리게 된다. (위키)

이름	수학 (x_1) 공부시간	영어 (x_2) 공부시간	합/불 (y)
야쓰오	2	3	0
오공	4	8	0
블리츠	6	2	1
페이커	8	4	1



Logistic Regression (개요)

❑ Linear Regression으로 Classification 을 해볼 수 있지 않을까?



이미지출처: 위키피디아

Logistic Regression (개요)

- ❑ y값이 0 or 1이 나와야 하는데..
  - ❑ 가능한 모든 X에 f(X)가 0 or 1값을 반환하는 함수를 어떻게 찾을 수 있을까?

H(X)=f(X) = $w_1 * x_1 + w_2 * x_2 + b$
= 0 or 1

$$Pr(Y_i = y_i | x_{1,i}, \ldots, x_{m,i}) = \begin{cases} p_i & \text{if } y_i = 1 \\ 1 - p_i & \text{if } y_i = 0 \end{cases}$$

이름	수학 (x_1) 공부시간	영어 (x_2) 공부시간	합/불 (y)
야쓰오	2	3	0
오공	4	8	0
블리츠	6	2	1
페이커	8	4	1

CONTENTS

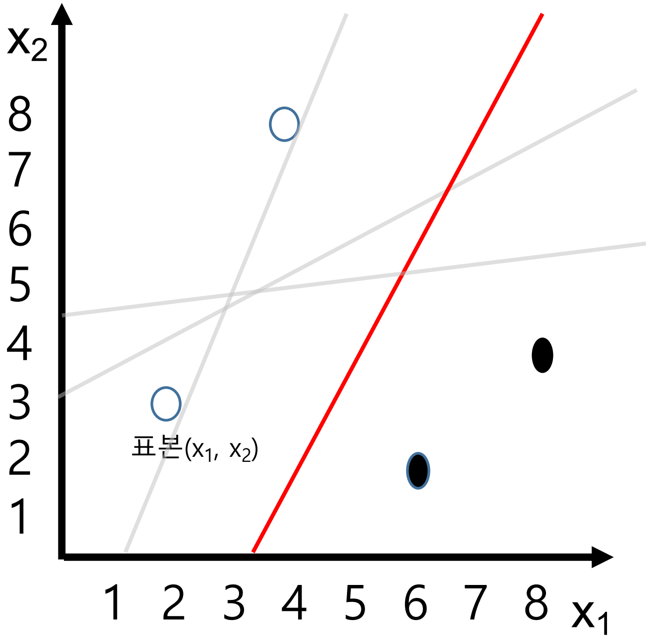
Sigmoid function

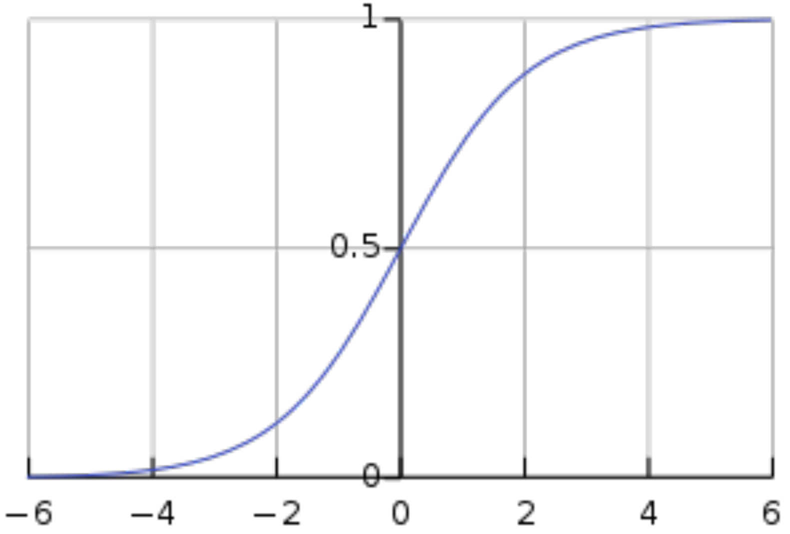
❑ 어떤 값이든 0~1사이로 변경해주는 Sigmoid 함수를 활용하자

$$S(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1}$$

$H(X)=f(X) = w_1 * x_1 + w_2 * x_2 + b$ → $S(H(X)) = S(w_1 * x_1 + w_2 * x_2 + b)$

$= 1/1+e^{-(w1*x1 + w2*x2 + b)}$

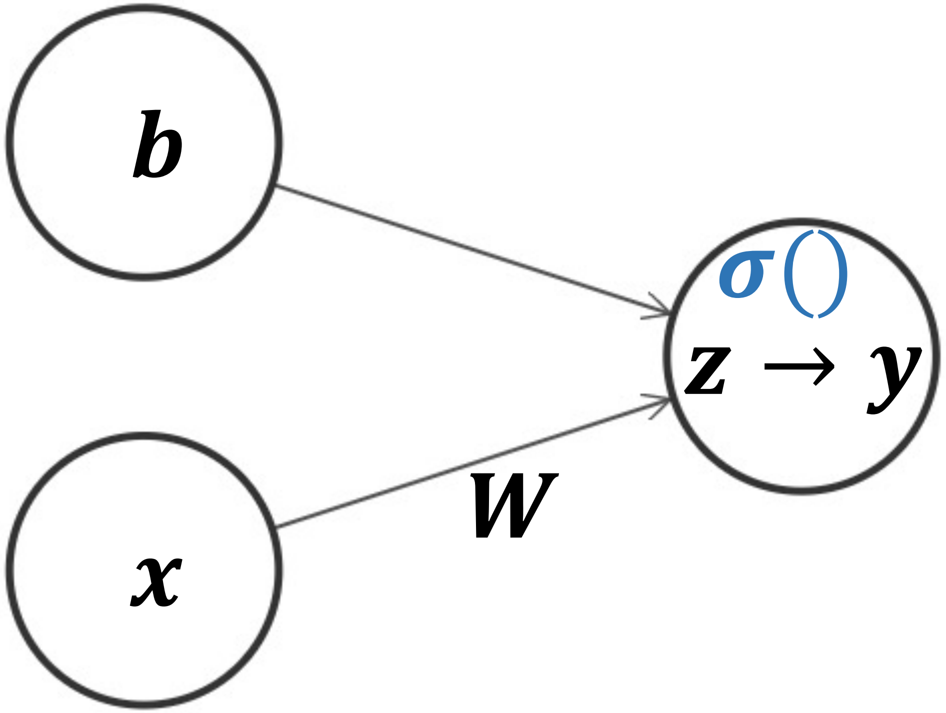


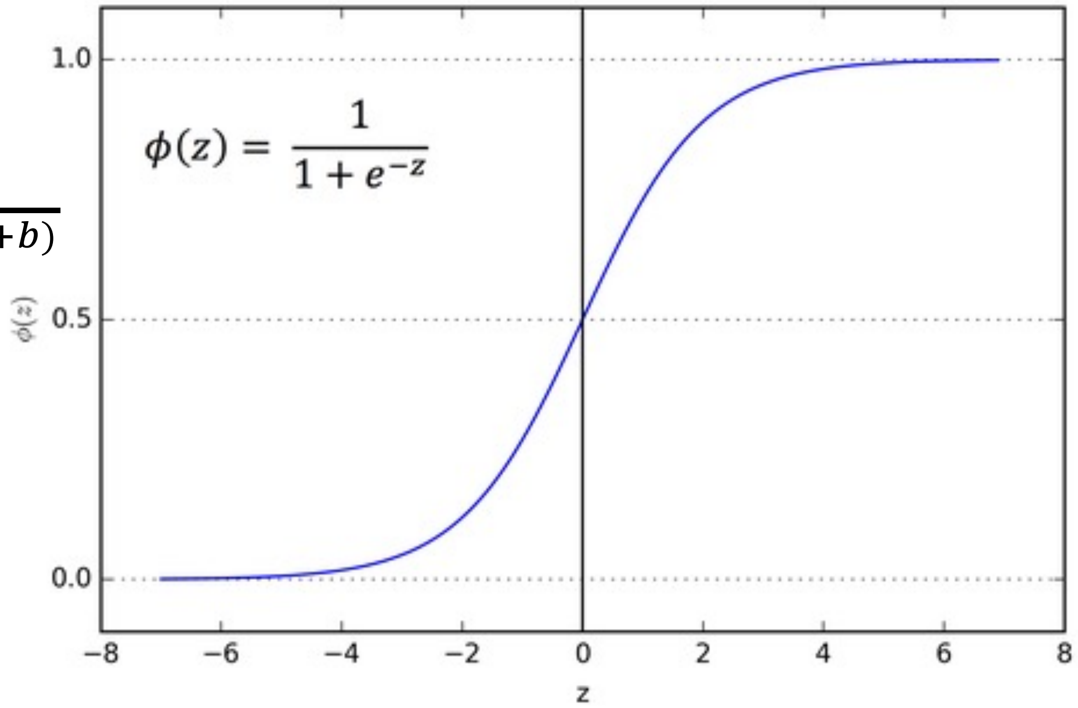


이미지출처: 위키피디아

❑ Sigmoid의 여러 표현 방식

- $z = Wx + b$
- $y = sigmoid(z) = \sigma(z) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(wx+b)}}$



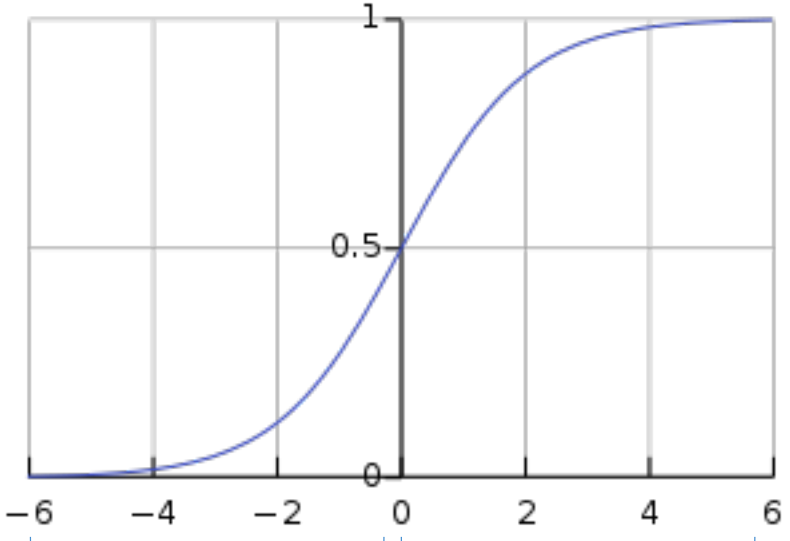


Sigmoid function

- ❑ 어떤 값이든 0~1사이로 변경해주는 Sigmoid 함수를 활용하자
  - ❑ S(X) 값이 0.5보다 크면 1, 작으면 0으로 가정해보자

$$S(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1}$$

$S(H(X)) = S(w_1 * x_1 + w_2 * x_2 + b)$
$\quad = 1/1+e^{-(w1*x1 + w2*x2 + b)}$



H(X) 값이 음수면
S(H(X))< 0.5

H(X) 값이 양수면
S(H(X))=>0.5

즉
H(X) =>0 이면 y=1로 예측
H(X) < 0 이면 y=0으로 예측

즉
$w_1 * x_1 + w_2 * x_2 + b$ =>0 이면 y=1로 예측
$w_1 * x_1 + w_2 * x_2 + b < 0$ 이면 y=0으로 예측

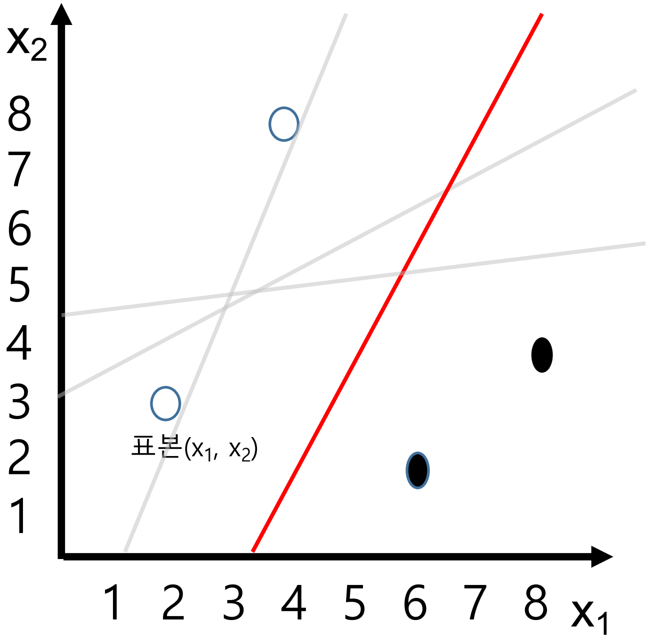
Sigmoid 데모: https://www.desmos.com/calculator/bgontvxotm?lang=ko

이미지출처: 위키피디아

Sigmoid function

❑ 어떤 값이든 0~1사이로 변경해주는 Sigmoid 함수를 활용하자

$H(X)=f(X) = w_1*x_1 + w_2*x_2 + b$



즉

$w_1*x_1 + w_2*x_2 + b$ =>0 이면 y=1로 예측

$w_1*x_1 + w_2*x_2 + b < 0$ 이면 y=0으로 예측

이 말은

$w_1*x_1 + w_2*x_2 + b = 0$ 인 선을 기준으로

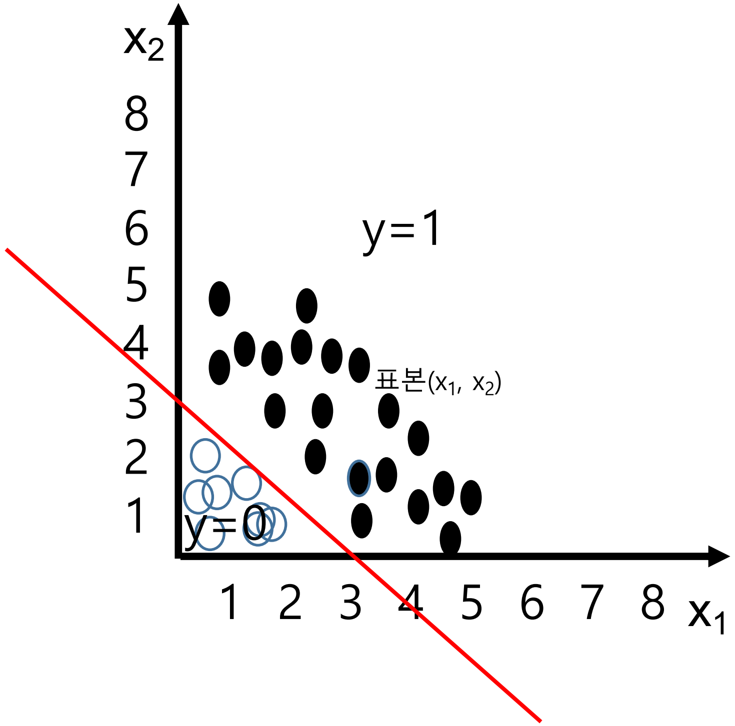
- 한쪽은 1
- 다른 한쪽은 0

우리는 이 선을 decision boundary라고 함

Logistic Regression (graphical 예제)

❑ 데이터 A에 대해 Linear decision boundary를 찾았다고 가정해 보자

$H(X)=f(X) = w_1*x_1 + w_2*x_2 + b$



w_1, w_2, b_1 을 1, 1, -3로 둔다면

$x_1 + x_2 - 3$ =>0 이면 y=1로 예측
$x_1 + x_2 - 3 < 0$ 이면 y=0으로 예측

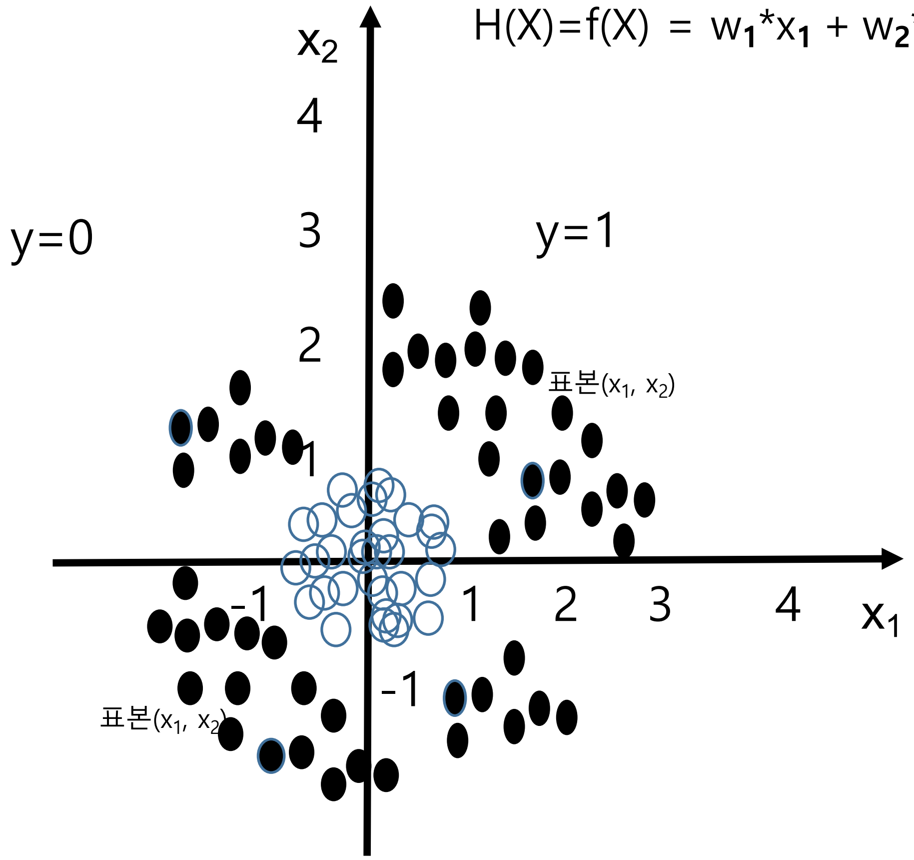
$x_1 + x_2$ =>3 이면 y=1로 예측
$x_1 + x_2 < 3$ 이면 y=0으로 예측

이 말은
$x_1 + x_2 - 3 = 0$ 인 선을 기준으로
- 한쪽은 1
- 다른 한쪽은 0

Logistic Regression (graphical 예제)

- Linear로 분류가 안될 수도 있겠지?
  - 실제 예제 Non-linear decision boundary를 보자

$H(X)=f(X) = w_1*x_1 + w_2*x_2 + w_3*x^2_1 + w_4*x^2_2 + b$



w_1, w_2, w_3, w_4, b 을 0, 0, 1, 1, -1로 둔다면

$x^2_1 + x^2_2 - 1 => 0$ 이면 y=1로 예측
$x^2_1 + x^2_2 - 1 < 0$ 이면 y=0으로 예측

$x^2_1 + x^2_2 => 1$ 이면 y=1로 예측
$x^2_1 + x^2_2 < 1$ 이면 y=0으로 예측

이 말은
$x^2_1 + x^2_2 - 1 = 0$ 인 원을 기준으로
- 한쪽은 1
- 다른 한쪽은 0

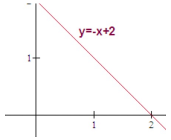
예제출처: Machine Learning (Stanford, Andrew NG)

CONTENTS

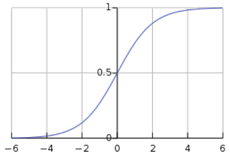
Logistic Regression의 Cost 함수

- Logistic Regression의 Cost 함수를 MSE로 해볼까?

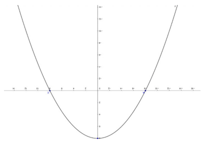
  - H(x) $= w_1 * x_1 + w_2 * x_2 + b$



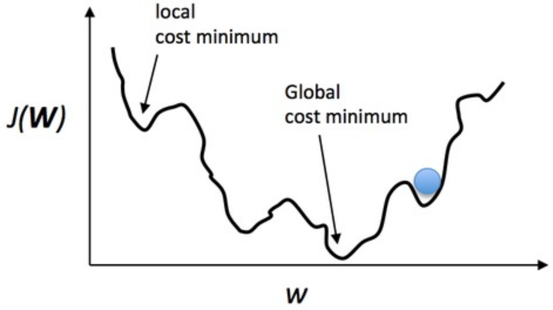
  - $$S(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1}$$



  - MSE $$cost(W,b) = \frac{1}{m}\sum_{i=1}^{m}(H(x^{(i)}) - y^{(i)})^2$$



  - 1차원 함수에 MSE를 하면 Cost값은 2차원 함수로 표현됨, 하지만 1차원 함수에 Sigmoid를 취하면??



Cost 함수에 log를 취해주면?
 - 주름이 펴진다 (convex)

Logistic Regression의 Cost 함수

- Linear Regression과 같이 MSE를 사용하지 않는 이유

Logistic Regression 에서는 Sigmoid 함수를 통과하기 때문에 비선형 함수가 출력됨.

Loss function을 최적화 하기 위해서 Gradient descent 등의 방법을 통하여 최소화해야 하는데 MSE의 형태 $(y - \hat{y})^2$을 보면 제곱이 되어있는 상태임.

비선형 함수를 제곱하게 되면 매우 복잡한 함수가 나올 수 있음

→ Local minima가 Global minima가 아닐 가능성이 높아짐

Logistic Regression의 Cost 함수

- 그렇다면 Logistic Classification의 Cost 함수는 어떻게 정의하면 될까?

  - Cost함수의 목적에 충실해지자!
    - 예측이 틀릴 경우 Cost = 매우 높음
    - 예측이 맞을 경우 Cost = 0

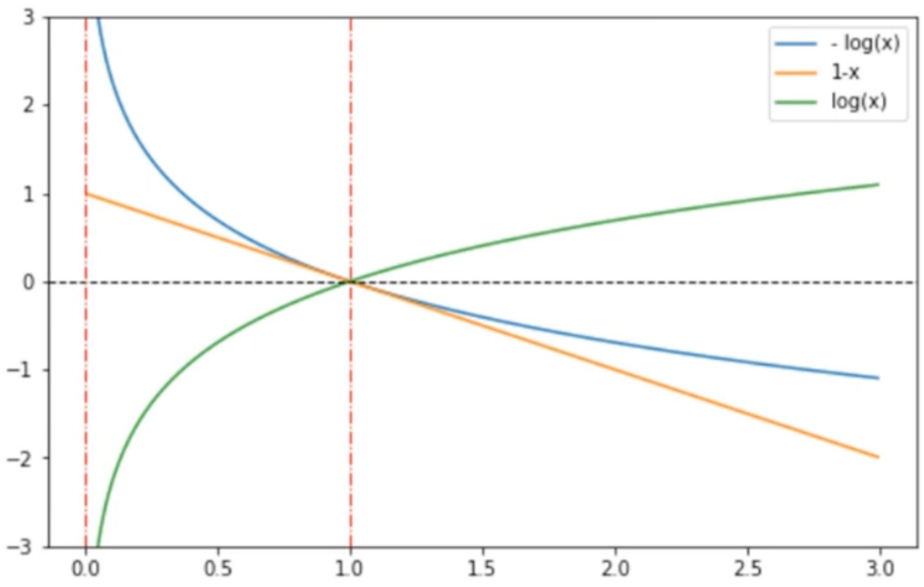
```
if: y==1 and S(x)==1: #정답과 예측이 일치
   return loss=0
if: y==0 and S(x)==0: #정답과 예측이 일치
   return loss=0
if: y==1 and S(x)==0: #정답과 예측이 불일치
   return loss=무한대
if: y==0 and S(x)==1: #정답과 예측이 불일치
   return loss=무한대
```

Logistic Regression의 Cost 함수

- 그렇다면 Logistic Classification의 Cost 함수는 어떻게 정의하면 될까?
  - Cost함수의 목적에 충실해지자!
    - 예측이 틀릴 경우 Cost = 매우 높음
    - 예측이 맞을 경우 Cost = 0

$h_\theta(x) = S(H(X))$,where θ = {W, b}로 간략화 하자!

$$\mathrm{cost}\big(h_\theta(x), y\big) = \begin{cases} -\log\big(h_\theta(x)\big) & \text{if } y = 1 \\ -\log\big(1 - h_\theta(x)\big) & \text{if } y = 0 \end{cases}$$




```
if: y==1 and S(x)==1: #정답과 예측이 일치
   return loss=0
if: y==0 and S(x)==0: #정답과 예측이 일치
   return loss=0
if: y==1 and S(x)==0: #정답과 예측이 불일치
   return loss=무한대
if: y==0 and S(x)==1: #정답과 예측이 불일치
   return loss=무한대
```

Logistic Regression의 Cost 함수 (Negative Log Likely-hood == Binary Cross Entropy)

- Y == 0, Y==1 두 경우의 손실 함수를 하나로 합쳐보자!

  - $Cost(W,b) = \frac{1}{m}\sum_{i=1}^{m} -ylog\left(h_\theta(x)\right) - (1-y)\log(1-h_\theta(x))$

$$J(w) = \sum_{i=1}^{m} y^{(i)} \log P(y=1) + (1-y^{(i)}) \log P(y=0)$$

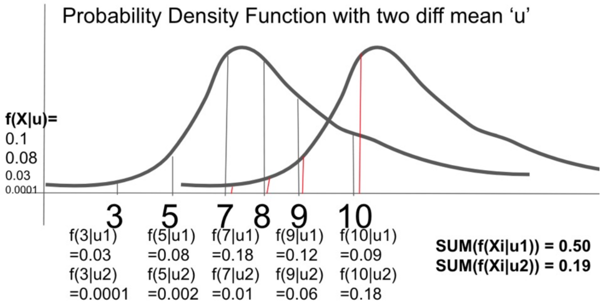
$$\left.\begin{aligned} \mathbb{P}(y=1|z) &= \sigma(z) = \frac{1}{1+e^{-z}} \\ \mathbb{P}(y=0|z) &= 1-\sigma(z) = \frac{1}{1+e^{z}} \end{aligned}\right\} \mathbb{P}(y|z) = \sigma(z)^y(1-\sigma(z))^{1-y}$$

Binary Cross Entropy

- $E(W,b) = -\Sigma_{i=1}^{n}\{t_i \log y_i + (1-t_i)\log(1-y_i)\}$
- Classification의 최종 출력 값 y는 sigmoid 함수에 의해 0~1사이 값을 가지는 확률적인 분류 모델이므로 다음과 같이 확률변수 C를 이용해 출력 값을 표현 가능

입력 X에 대해 출력 값이 1일 확률을 y로 정의할 경우

- $p(C=1|x) = y = sigmoid(Wx+b),\ where\ y = 0{\sim}1$
- $p(C=0|x) = 1-y\,,\quad p(C=t|x) = y^t(1-y)^{1-t}\,, t \in \{0,1\}$



가능도 (**L**ikelyhood) : 데이터를 관찰함으로써 이 데이터가 추출되었을 것으로 생각되는 분포의 특성 (모수=파라미터) 값을 추정

$$\boldsymbol{L}(W,b) = \Pi_{i=1}^{n}\, p(C=t_i|x_i) = \Pi_{i=1}^{n} y_i^{t_i}(1-y_i)^{1-t_i} \qquad \cdots\text{(MLE)}$$

입력 x에 대해 정답 t가 발생될 확률을 나타낸 함수, 확률은 독립적이므로 각각 입력데이터의 발생 확률을 곱해서 우도함수로 나타냄

$$E(W,b) = -\log L(W,b) = -\Sigma_{i=1}^{n}\{t_i \log y_i + (1-t_i)\log(1-y_i)\}$$

함수 값이 최대가 되는 것을 알기 위해 w, b값을 L함수로 편미분 해야하는데 곱하기는 미분이 불편하므로 양변에 log를 취하여 덧셈 형태로 바꿈

Logistic Regression의 Cost 함수 정리

- Linear Regression의 경우 MSE를 활용하여 파라미터 최적화

- Logistic Regression의 경우 BCE를 활용하여 파라미터 최적화
- BCE의 경우 확률/통계, 정보 이론과 밀접한 관련

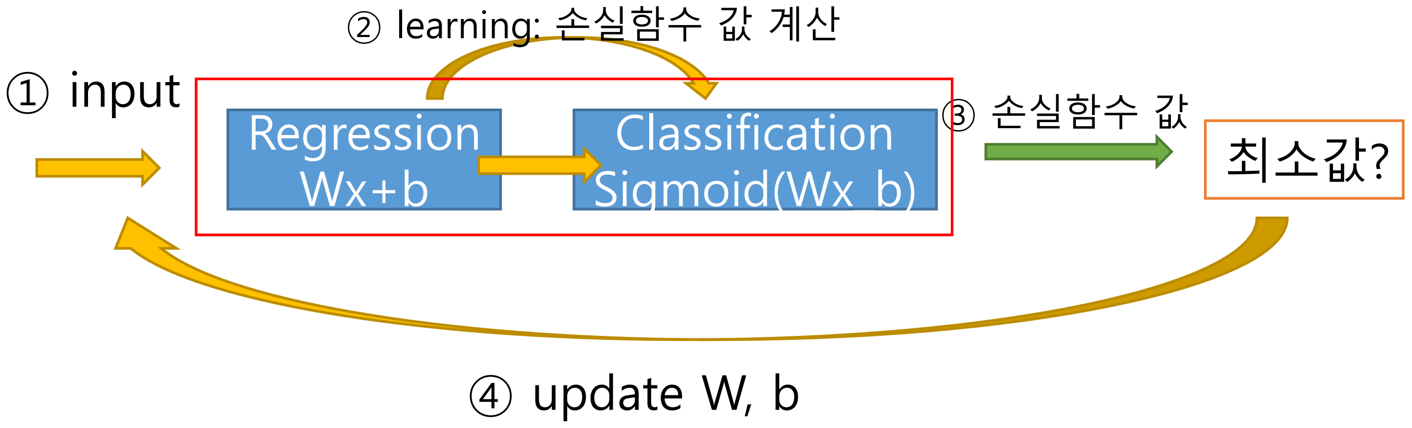
(유도과정에서 또한 MLE가 사용되며 BCE의 이름에서 Entropy와 연관됨)

Logistic Regression의 Gradient Descent

- Cost값을 기반으로 Gradient Decent 알고리즘을 이용해 최적의 파리미터 W값 을 찾자!
  - $H(x) = w_1 * x_1 + w_2 * x_2 + b$
  - $Cost(W,b) = \frac{1}{m}\sum_{i=1}^{m} -y\log\left(h_\theta(x)\right) - (1-y)\log(1-h_\theta(x))$
  - $\underset{W,b}{\text{minimize}}\; cost(W,b)$ $\qquad W := W - \alpha\frac{\partial}{\partial W}cost(W)$
  - $\frac{\partial}{\partial\theta}Cost(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(Sh_\theta\left(x^i\right) - y^i\right)x^i$

공부시간	합/불
9	불합격
14	불합격
21	불합격
27	합격
32	합격
37	합격

학습 데이터



Logistic Regression 수식 정리!

- Logistic Classification을 정리하면!
  - $H(x) = w_1 * x_1 + w_2 * x_2 + b$
  - $S(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1}$
  - $S(H(X)) = \frac{1}{1+e^{-(w1*x1\ +\ w2*x2\ +\ b)}}$
  - $Cost(W,b) = \frac{1}{m}\sum_{i=1}^{m} -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$
  - $\underset{W,b}{\text{minimize}}\, cost(W,b)$ $\quad W := W - \alpha\frac{\partial}{\partial W}cost(W)$

Logistic Regression 정리!

- 독립 변수의 선형 결합을 이용하여 사건의 발생 가능성을 예측하는데 사용되는 통계 기법
- Logistic Regression 모델은 Linear Regression 모델에 Sigmoid 함수 적용
- Logistic Regression은 회귀를 사용하여 데이터가 어떤 class에 속할 0에서 1사이의 값으로 예측하고 이를 통해서 class를 분류함.
- MLE(Maximum Likelihood Estimation)을 사용하여 학습하기 때문에 충분한 표본이 필요함.

APPENDIX

Logistic Regression 미분과정 정리!

Appendix: Partial derivative of $J(\theta)$

우선 logistic funtion의 미분을 구해두자.

$$\sigma(x)' = \left(\frac{1}{1+e^{-x}}\right)' = \frac{-(1+e^{-x})'}{(1+e^{-x})^2} = \frac{-1' - (e^{-x})'}{(1+e^{-x})^2} = \frac{0-(-x)'(e^{-x})}{(1+e^{-x})^2} = \frac{-(-1)(e^{-x})}{(1+e^{-x})^2} = \frac{e^{-x}}{(1+e^{-x})^2}$$
$$= \left(\frac{1}{1+e^{-x}}\right)\left(\frac{e^{-x}}{1+e^{-x}}\right) = \sigma(x)\left(\frac{+1-1+e^{-x}}{1+e^{-x}}\right) = \sigma(x)\left(\frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}\right) = \sigma(x)(1-\sigma(x))$$

이제 편미분을 구해보자.

$$\begin{aligned}
\frac{\partial}{\partial\theta_j}J(\theta) &= \frac{\partial}{\partial\theta_j}\frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right] \\
&= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\frac{\partial}{\partial\theta_j}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\frac{\partial}{\partial\theta_j}log(1-h_\theta(x^{(i)}))\right] \\
&= -\frac{1}{m}\sum_{i=1}^{m}\left[\frac{y^{(i)}\frac{\partial}{\partial\theta_j}h_\theta(x^{(i)})}{h_\theta(x^{(i)})} + \frac{(1-y^{(i)})\frac{\partial}{\partial\theta_j}(1-h_\theta(x^{(i)}))}{1-h_\theta(x^{(i)})}\right] \\
&= -\frac{1}{m}\sum_{i=1}^{m}\left[\frac{y^{(i)}\frac{\partial}{\partial\theta_j}\sigma(\theta^T x^{(i)})}{h_\theta(x^{(i)})} + \frac{(1-y^{(i)})\frac{\partial}{\partial\theta_j}(1-\sigma(\theta^T x^{(i)}))}{1-h_\theta(x^{(i)})}\right] \\
&= -\frac{1}{m}\sum_{i=1}^{m}\left[\frac{y^{(i)}\sigma(\theta^T x^{(i)})(1-\sigma(\theta^T x^{(i)}))\frac{\partial}{\partial\theta_j}\theta^T x^{(i)}}{h_\theta(x^{(i)})} + \frac{-(1-y^{(i)})\sigma(\theta^T x^{(i)})(1-\sigma(\theta^T x^{(i)}))\frac{\partial}{\partial\theta_j}\theta^T x^{(i)}}{1-h_\theta(x^{(i)})}\right] \\
&= -\frac{1}{m}\sum_{i=1}^{m}\left[\frac{y^{(i)}h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))\frac{\partial}{\partial\theta_j}\theta^T x^{(i)}}{h_\theta(x^{(i)})} - \frac{(1-y^{(i)})h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))\frac{\partial}{\partial\theta_j}\theta^T x^{(i)}}{1-h_\theta(x^{(i)})}\right] \\
&= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}(1-h_\theta(x^{(i)}))x_j^{(i)} - (1-y^{(i)})h_\theta(x^{(i)})x_j^{(i)}\right] \\
&= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}(1-h_\theta(x^{(i)})) - (1-y^{(i)})h_\theta(x^{(i)})\right]x_j^{(i)} \\
&= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} - y^{(i)}h_\theta(x^{(i)}) - h_\theta(x^{(i)}) + y^{(i)}h_\theta(x^{(i)})\right]x_j^{(i)} \\
&= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} - h_\theta(x^{(i)})\right]x_j^{(i)} \\
&= \frac{1}{m}\sum_{i=1}^{m}\left[h_\theta(x^{(i)}) - y^{(i)}\right]x_j^{(i)}
\end{aligned}$$

출처 https://wikidocs.net/4289

Sigmoid function

공부 시간	합/불
9	불합격
37	합격

학습 데이터

→ Regression → Classification →

공부 시간	합/불
9	??
37	??

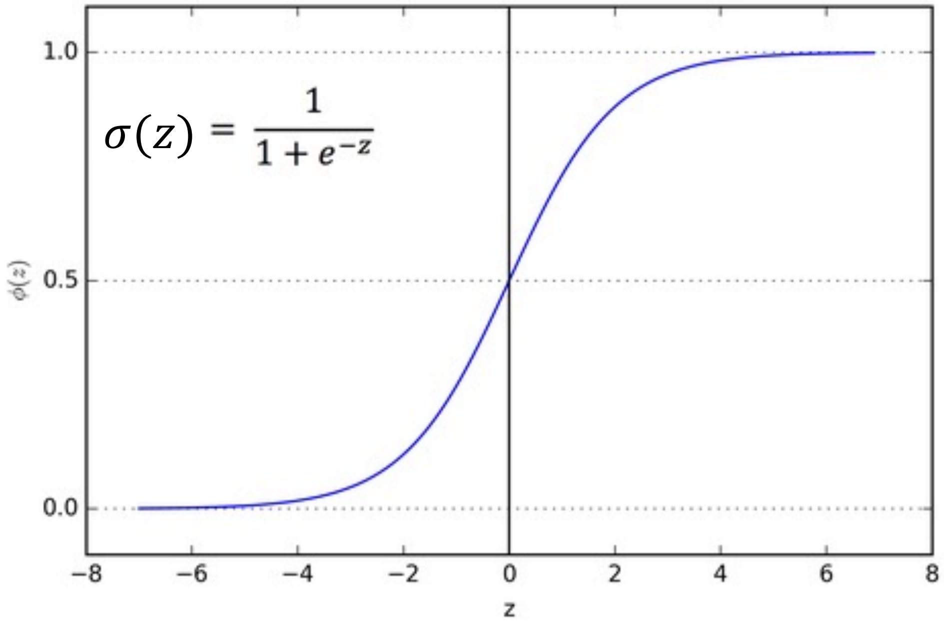
예측

- $sigmoid(x) = \sigma(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x}$

- 출력 값이 0과 1을 가져야 하는 즉, 분류 시스템에서 0~1사이의 값을 갖는 sigmoid함수를 이용함.

Cf) $y > 0.5 \rightarrow y = 1,\quad y \leq 0.5 \rightarrow y = 0$

Sigmoid function



- Sigmoid function, Logistic function, Squashing function으로 불림.
- Output : 0~1 사이 범위
- 단조증가 함수이기 때문에 $f(x)$와 $\sigma\big(f(x)\big)$는 같은 x에서 최소값을 가짐.
- 미분결과를 sigmoid로 표현 가능, $\frac{d}{dx}\sigma(x) = \sigma(x)\big(1-\sigma(x)\big)$

- 승산(Odds)

성공 확률을 p라고 할 때, 성공 확률 / 실패 확률

$$Odds = \frac{p}{1-p}$$

$$p = 1 \ \rightarrow odds = \infty$$

$$p = 0 \ \rightarrow odds = 0$$

Sigmoid function

- A와 B의 범주를 갖는 이진분류를 생각해보자.

$$E(y) = \pi(X = x) = P(Y = A|X = x) = 1 - P(Y = B|X = x)$$ 이면

$$\pi(X = x) = \frac{1}{1+e^{-(wx+b)}}$$ 이므로, $0 \leq \pi(X = x) \leq 1$ 이고

$$odds = \frac{\pi(X=x)}{1-\pi(X=x)}$$ 이다.

($\pi(X = x)$: A범주일 확률, $1 - \pi(X = x)$:B범주일 확률)

$$\log(odds) = \log\left(\frac{\pi(X=x)}{1-\pi(X=x)}\right) = \log\left(\frac{\frac{1}{1+e^{-(wx+b)}}}{1-\frac{1}{1+e^{-(wx+b)}}}\right) = wx + b$$

→ 선형결합의 형태로 변함


감사합니다.