3. Classification

다루면. 이변수 → regression!?
생각 보다. 어렵다.
regression ⇐ new feature
(사람의 개념).

분류

model → predict → 확률

0.5↑ → True
0.5↓ → False.

복숭아 수확 기계

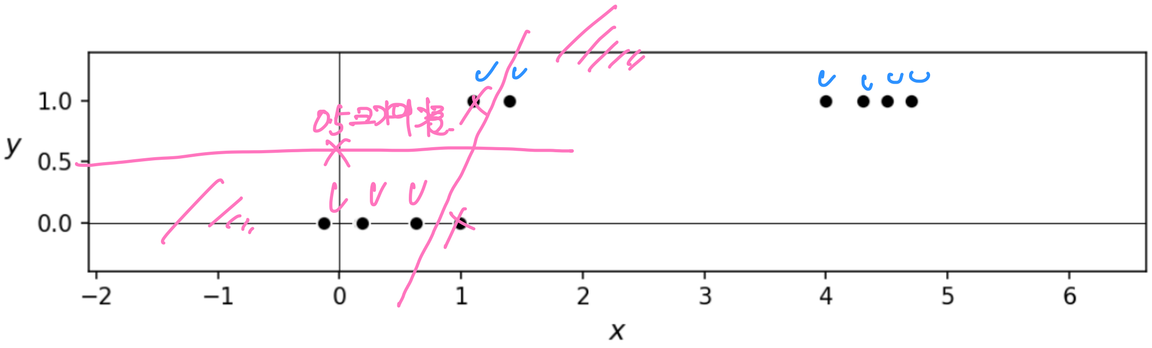
https://cdn.ppomppu.co.kr/zboard/data3/2021/0103/20210103221124_ebzlkcra.mp4

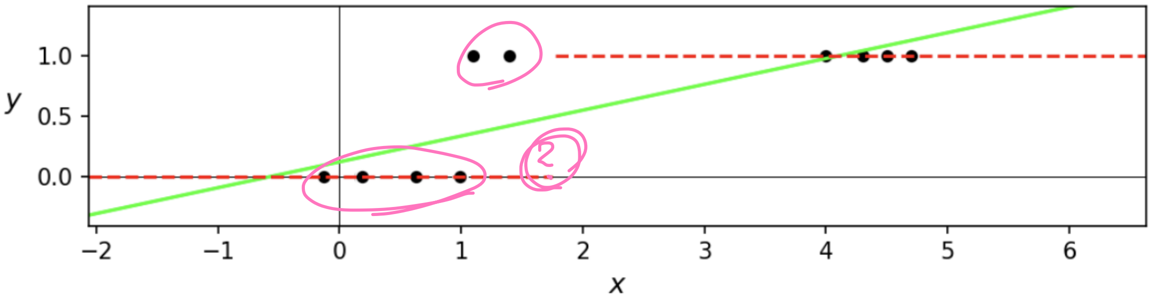
- http://m.ppomppu.co.kr/new/bbs_view.php?id=humor&no=423793
- 익은 복숭아 vs 안익은 복숭아
- 크기, 물렁도, 색깔, ...

차원에 따른 분류 문제

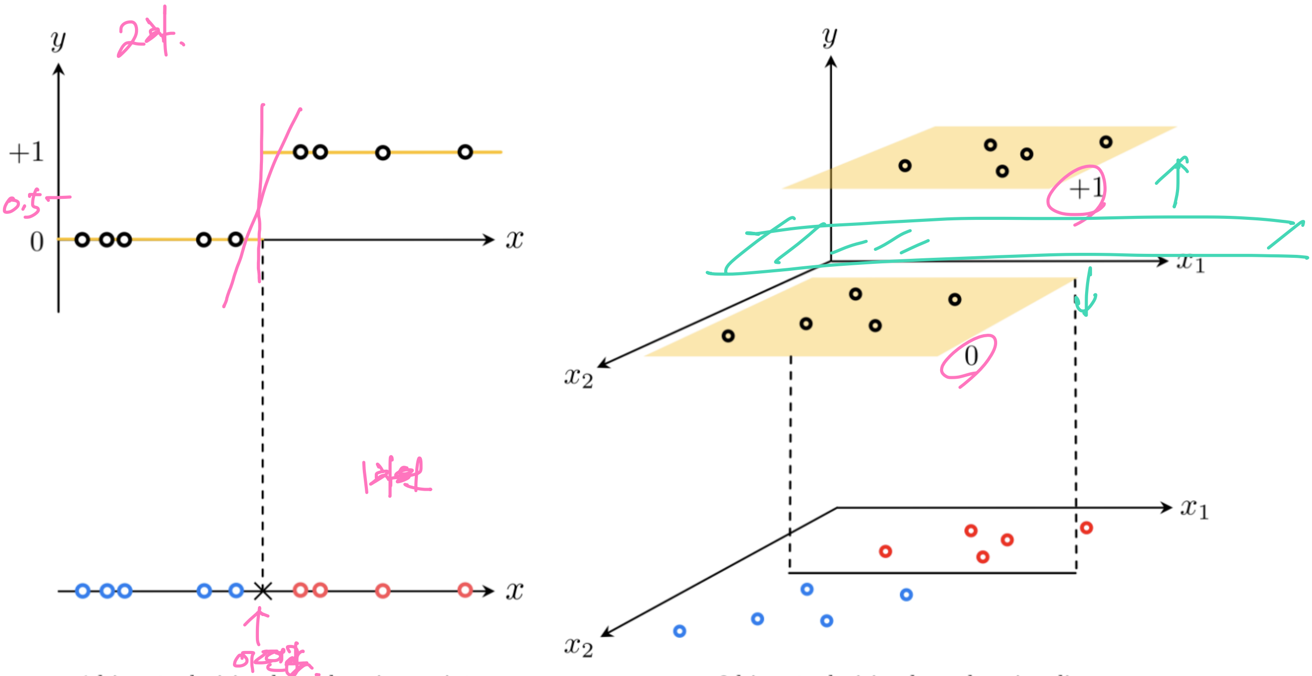
직선을 이용한 분류 문제 해결

예. 크기





차원에 따른 분류문제 해결법



1d input: decision boundary is a point

2d input: decision boundary is a line

- **Bottom Step**: 대부분의 레이블 값이 $y_p = 0$인 점들을 포함합니다.
- **Top Step**: 대부분의 레이블 값이 $y_p = +1$인 점들을 포함합니다.
- **분리**
    - $N = 1$일 때는 점으로, $N = 2$일 때는 선으로, $N > 2$일 때는 초평면 (hyperplane)으로 대부분 분리된다.

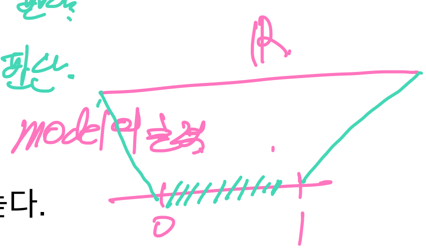
Classification with Probability

- 확실성을 가지고 예측할 수 없는 많은 상황에서, 어떤 사건이 발생할 확률을 예측하는 데 유용하기 위해서는 어떻게 해야 할까?
- 예시: 암 예측, 주식/비트코인 가격 예측, 날씨 예측 등.

기준: 0.5 $\quad$ model$(x_1) > 0.5 \Rightarrow$ True 판단

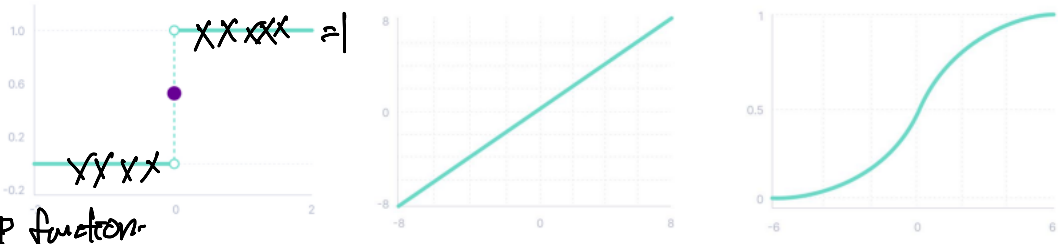
model$(x_2) \le 0.5 \Rightarrow$ False 판단.

Continuous Output



- 출력 y 는 0과 1 사이에서 연속적으로 변화한다.
- y 가 1에 가까울수록 어떤 사건이 발생할 가능성이 높다.

- **Soft Binary Classification):**
  - 소프트 레이블(확률, 0 ~ 1)을 반환한다.
  - Probability Regression: 출력을 [0, 1] 범위로 제한한다.



Step function

<classification>   <linear regression (unbounded)>   **<logistic regression (bounded)>**

Function1 : Step function

결과 0.5보다큼 True

model(Cvalue) ⇒ 2.3 ⇒ 1

model(Cvalue) ⇒ 0.3 ⇒ 0

0.5보다작다. False

$$\text{step}(x) = \begin{cases} 1 & \text{if } x \geq 0.5 \\ 0 & \text{if } x < 0.5 \end{cases}.$$

model = $w_0 + xw_1$

model(x_0) = $w_0 + x_0 w_1$

- **선형 경계**: 두 클래스 간의 경계는 선형이며, 모든 점 x에 대해 $w_0 + xw_1 = 0.5$로 정의됩니다.

  경계.

- N차원 입력에 대한 모델
  - N차원 입력에 대한 선형 모델: 일반적인 N차원 입력을 위한 선형 모델은 다음과 같이 표현된다.

    $x = [1 \ x_1 \ x_2 \ \cdots \ x_N]$ $w = [w_0 \ w_1 \ w_2 \ \cdots \ w_N]$

    - $\cancel{\mathbf{x}^T\mathbf{w}} = w0 + x_1 w_1 + x_2 w_2 + \cdots + x_N w_N$ $w^T = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_N \end{bmatrix}$

      xw^T
  - step function과의 결합: 해당 선형 결합에 단계 함수를 적용하면 $\text{step}\left(\mathbf{x}^T\mathbf{w}\right)$ 와 같다.

$x \cdot w^T = \rightarrow \downarrow = w_0 + x_1 w_1 + \cdots$

파라미터 조정

- 파라미터 조정의 목표
  - 적합한 가중치 설정
  - 분류기는 최종적으로 레이블이 $+1$인 점은 $\mathbf{x}^T\mathbf{w} > 0.5$*인 양의 영역, 레이블이* 0*인 점은* $\mathbf{x}^T\mathbf{w} < 0.5$인 음의 영역에 위치하도록 분류해야 한다.

- 이를 통해 각 점의 step $\left(\mathbf{x}^T\mathbf{w}\right)$ 에 대한 예측값이 해당 점의 Label 값과 일치하도록 한다.

$w_0 + w_1x_1 + w_2x_2 =$ model(분석 1)

- Cost function

$$E(\mathbf{w}) = \frac{1}{P}\sum_{p=1}^{P}\left(\text{step}\left(\mathbf{x}^T\mathbf{w}\right) - y_p\right)^2$$

y_p : 있으면 1, 없으면 0

문제점

- 최소 제곱 비용의 이산성
  - E(w)의 결과는 정수 값만을 취한다. 따라서 미분값이 모두 0 이다. 따라서 경사하강법 적용 불가

미분X → 최적화X.

- Step functiondml의 불연속성
  - Step function 은 불연속 함수 있다. 따라서 Cost function도 불연속 적인 성질을 갖고 있다.

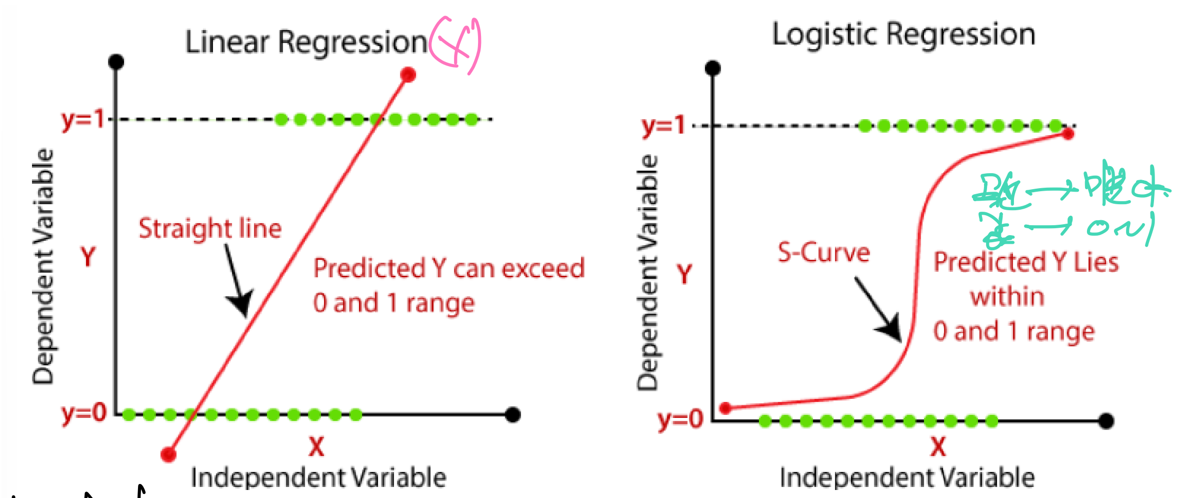
E(W)

연속, 기울기, 확률

Function2 : Logistic Function, Sigmoid Function

- 선형 함수 $s = w^Tx + b$ 를 0과 1 사이로 제한하는 함수.
- Sigmoid함수 $\sigma(s)$는$\sigma(s) = \frac{1}{1+e^{-s}}$ 로 정의된다.
- 입력 s 는 $[-\infty, +\infty]$, 출력은 $[0, 1]$ 범위.

$s \Rightarrow \frac{1}{1+e^{-s}}$



Step function
↳ 0, 1 아닌 False True

작아지면 0
커지면 1.

Sigmoid function
$0 < \frac{1}{1+e^{-s}} < 1$ → 확률을 나타낸다.

$\sigma(s) = \frac{1}{1+e^{-s}}$ $\lim_{s\to-\infty} \frac{1}{1+e^{-s}} = 0$

$\lim_{s\to\infty} \frac{1}{1+e^{-s}} = \frac{1}{1+0} = 1$

odd → logit → Sigmoid

↓

Softmax

선형 회귀의 비용 함수

선형 회귀에서는 최소 제곱 오차(LSE, Least Squares Error) 기준을 사용합니다. 이는 모델이 예측한 값과 실제 값 사이의 차이의 제곱합으로 정의됩니다. 선형 회귀의 비용 함수는 다음과 같이 표현됩니다:

$$J_{\text{linear}}(\mathbf{w}) = \frac{1}{2P}\sum_{p=1}^{P}(y_p - \mathbf{x}_p^T\mathbf{w})^2$$

로지스틱 회귀와 최소 제곱 오차의 문제점

- 로지스틱 회귀에 선형 회귀와 같은 LSE 기준을 그대로 적용하면, 비용 함수는 다음과 같다

회귀결과 → Sigmoid 적용 → MSE 적용

↓

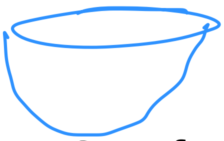
non-convex.

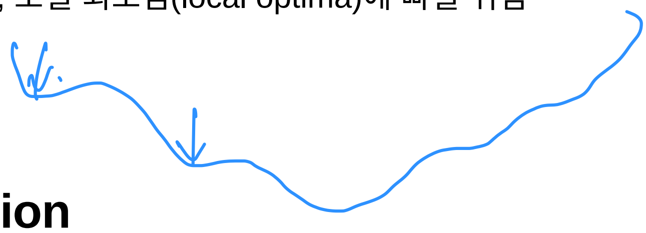
Sigmoid

$$J_{\text{logistic}}(\mathbf{w}) = \frac{1}{2P}\sum_{p=1}^{P}(y_p - \sigma(\mathbf{x}_p^T\mathbf{w}))^2$$

$$\frac{1}{1+e^{-x_p^T w}}$$

- 비용 함수는 비볼록(non-convex) 형태가 되어, 로컬 최소점(local optima)에 빠질 위험이 있다.

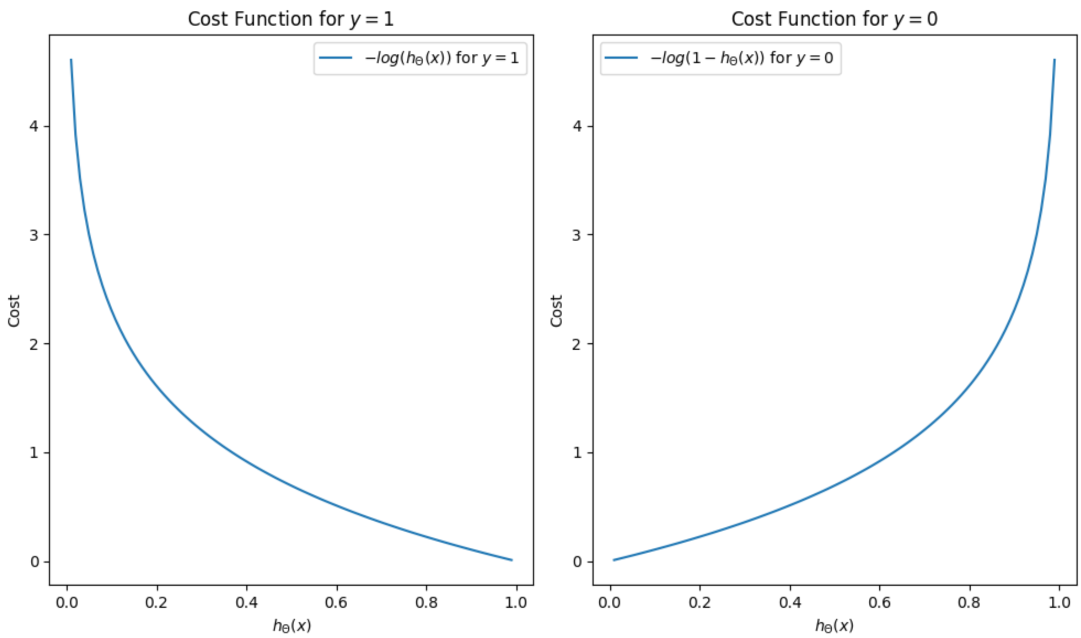




Logistic regression Costfunction

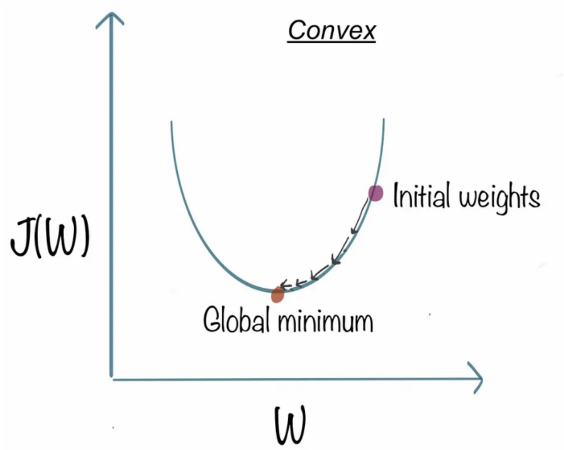
- **정의**: Log Error의 Cost function는 출력은 y_p가 1 또는 0일 때 다음과 같이 정의됩니다
  - y_p : p 에서의 y 값
  - $h_w(\mathbf{x}_p) = \sigma(\mathbf{x}_p^T w) = \frac{1}{1+e^{-\mathbf{x}_p^T w}}$

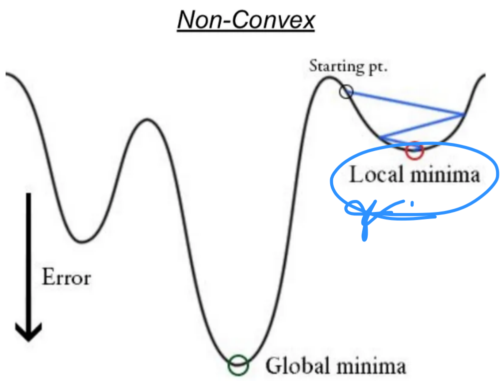
$$\text{cost}(h_w(\mathbf{x}_p), y_p) = \begin{cases} -\log(h_w(\mathbf{x}_p)) & \text{if } y_p = 1 \\ -\log(1 - h_w(\mathbf{x}_p)) & \text{if } y_p = 0 \end{cases}$$

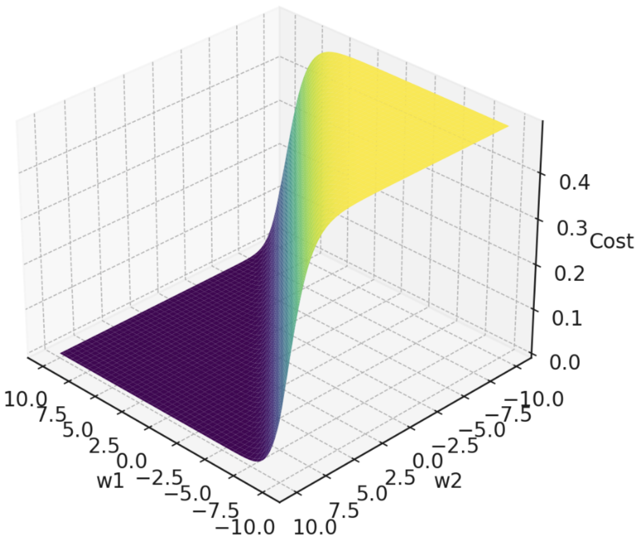


- Simplification form Cross entropy : Loss function

$$\mathrm{cost}(h_w(\mathbf{x}_p), y_p) = -y_p \log(h_w(\mathbf{x}_p)) - (1 - y_p)\log(1 - h_w(\mathbf{x}_p))$$







Logistic regression Costfunction

- 로지스틱 회귀에서는 교차 엔트로피 비용 함수를 사용한다.

Cross entropy

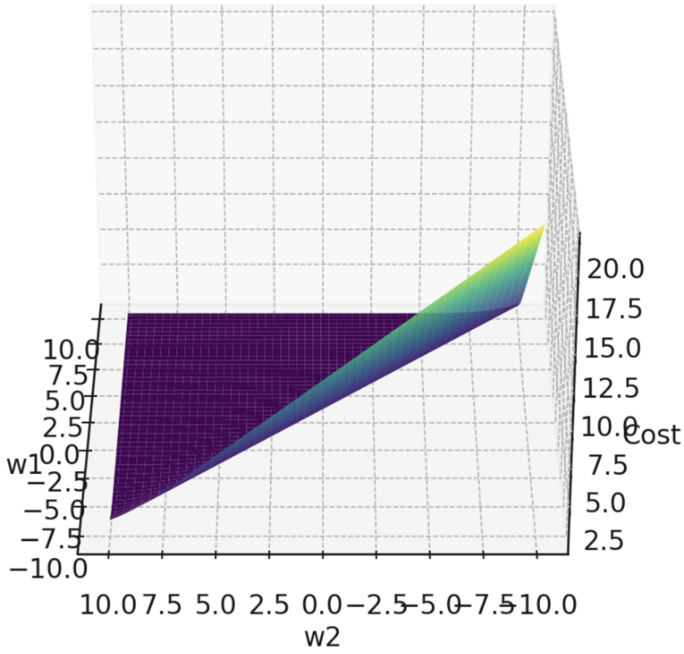
$$J_{\text{logistic}}(\mathbf{w}) = -\frac{1}{P}\sum_{p=1}^{P}\left[y_p \log(\sigma(\mathbf{x}_p^T\mathbf{w})) + (1-y_p)\log(1-\sigma(\mathbf{x}_p^T\mathbf{w}))\right]$$

원래 출력되는 확률

여사건

- 특징
  - True $y_p = 1$일 때, $\sigma(\mathbf{x}_p^T\mathbf{w})$가 1에 가까울수록 비용은 감소한다
  - False $y_p = 0$일 때, $\sigma(\mathbf{x}_p^T\mathbf{w})$가 0에 가까울수록 비용은 감소한다
  - 최대 우도 추정(Maximum Likelihood Estimation) 기준을 따른다
  - 볼록 함수(Convex)로, 전역 최소점(global minimum)을 찾는 것이 가능하다

$y_p=1$, $\sigma(x_p^T w)=0$
$y_p=0$, $\sigma(x_p^T w)=1$
인 경우에는 cost값이 증가한다

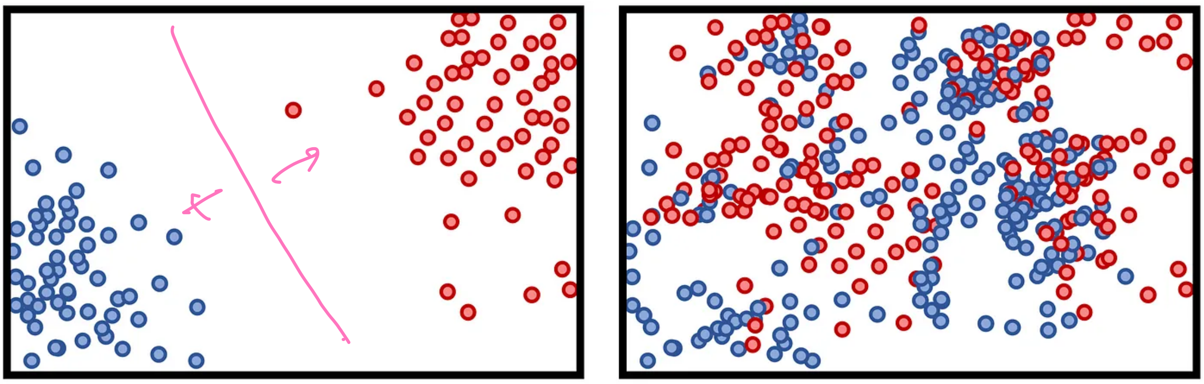

w1
w2
cost
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
20.0
17.5
15.0
12.5
10.0

예측(Prediction)

sigmoid.

- **클래스 선택**: 더 높은 확률을 가진 클래스가 선택된다.
  - 양성 Class : $P(y_i = 1|x_i) = \theta(w^T x_i + b)$
  - 음성 Class : $P(y_i = -1|x_i) = 1 - \theta(w^T x_i + b)$

교차 엔트로피 비용 함수



Low Entropy　　　　High Entropy

- **정의**: 교차 엔트로피 비용 함수는 모든 데이터 포인트에 대한 로그 오류 비용의 평균으로 정의된다.

$$J(\mathbf{w}) = \frac{1}{P} \sum_{p=1}^{P} cost_p(\mathbf{w})$$

- **등가 표현**: 로그 오류를 하나의 식으로 결합한 형태

$$J(\mathbf{w}) = -\frac{1}{P} \sum_{p=1}^{P} \left[y_p \log\left(\sigma\left(\mathbf{x}_p^T \mathbf{w}\right)\right) + (1 - y_p) \log\left(1 - \sigma\left(\mathbf{x}_p^T \mathbf{w}\right)\right) \right]$$

- **가중치 조정**: 이 비용 함수는 가중치 $\mathbf{w}$가 적절하게 조정되었을 때 최소값에 도달합니다.
- **최적화**: 로지스틱 회귀에서의 최적화에 적합하며, 이진 분류 문제에 유용합니다.

* linear regression ⇒ regression.

logistic regression ⇒ classification.

sigmoid.
cross-entropy

Cost function

Model Evaluate