선형분류(Linear Classification)

선형회귀(Linear Regression) → $[-\infty, \infty]$

×

로지스틱 회귀(Logistic Regression) = 시그모이드 함수(Sigmoid function)

$\Downarrow$

선형분류(Linear Classification) → $0 \, \mathrm{or} \, 1$

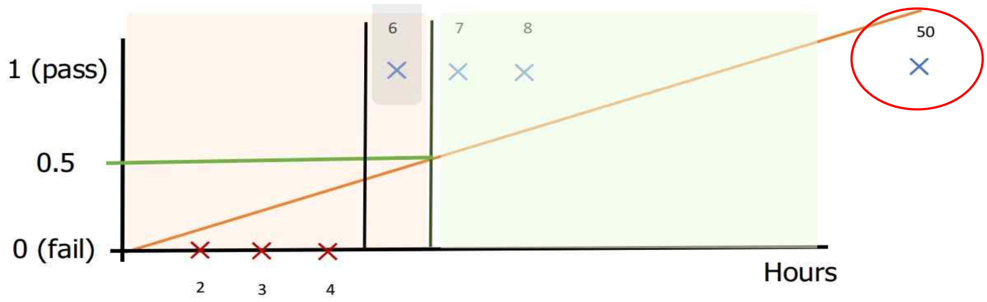
: 이진분류(Binary Classification) ← Sigmoid function

& 다중분류(Multiclass Classification) ← Softmax function

이진분류(Binary Classification)

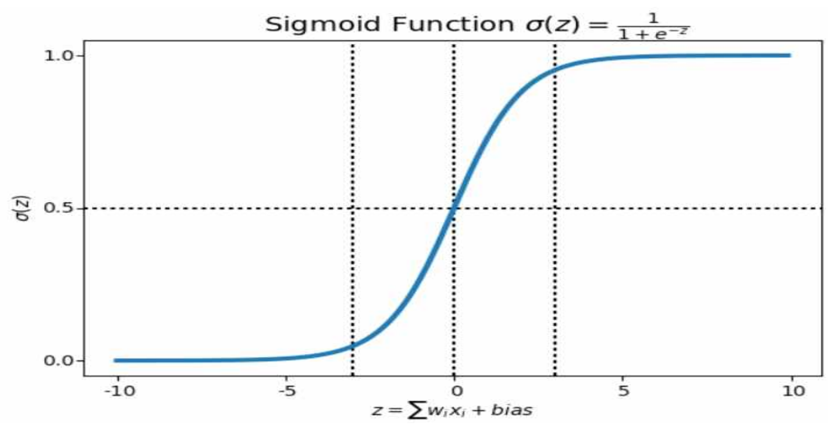
Binary Label Encoding → 0 or 1

Why 선형회귀는 분류문제를 풀지 못할까?



Sol) 가설 $H(x)$를 0과 1사이의 값으로 만들자 using Sigmoid function

Sigmoid function: $\sigma(Z) = \dfrac{1}{1+e^{-Z}}$



Logistic Hypothesis: $Z = H(X) = W^T X \Rightarrow H(X) = \frac{1}{1+e^{-W^T X}}$

BUT. Cost function: $cost(W,b) = \frac{1}{m}\sum_{i=1}^{m}(H(x^{(i)}) - y^{(i)})^2$에 적용해보면

$H(x) = Wx + b$

: convex function으로 global minimum이 한 개이지만

$H(X) = \frac{1}{1+e^{-W^T X}}$

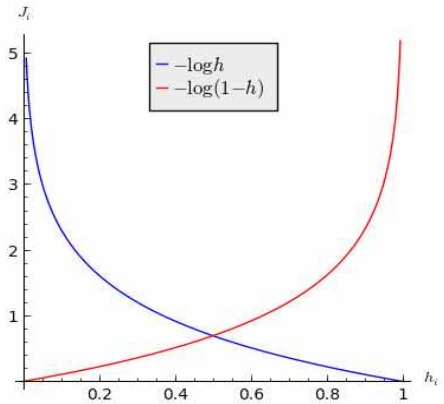
: non-convex function으로 local minimum을 여러 개 갖는다

가설이 변경되었으므로 Cost도 재설계해야한다

→ 예측값과 정답이 얼마나 가까운지

정답에 가까울수록 cost는 작고 정답에 멀수록 cost는 크게

H(x)는 0 or 1이므로 H(x) 〈 0.5 → 0, H(x) 〉 0.5 → 1



Cross-entropy:

$$c(H(X),y) = \begin{cases} -\log(H(X)) & : y=1 \\ -\log(1-H(X)) & : y=0 \end{cases}$$

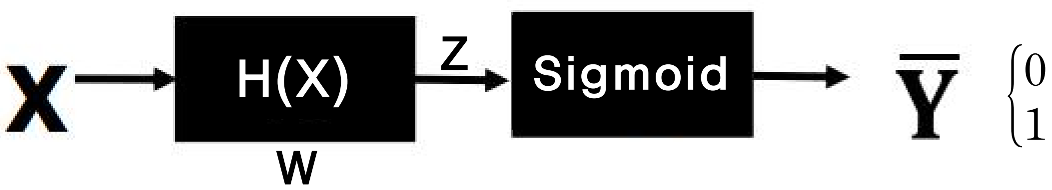
하나의 식으로 표현하면

$$c(H(X),y) = -y\log(H(x)) - (1-y)\log(1-H(x))$$

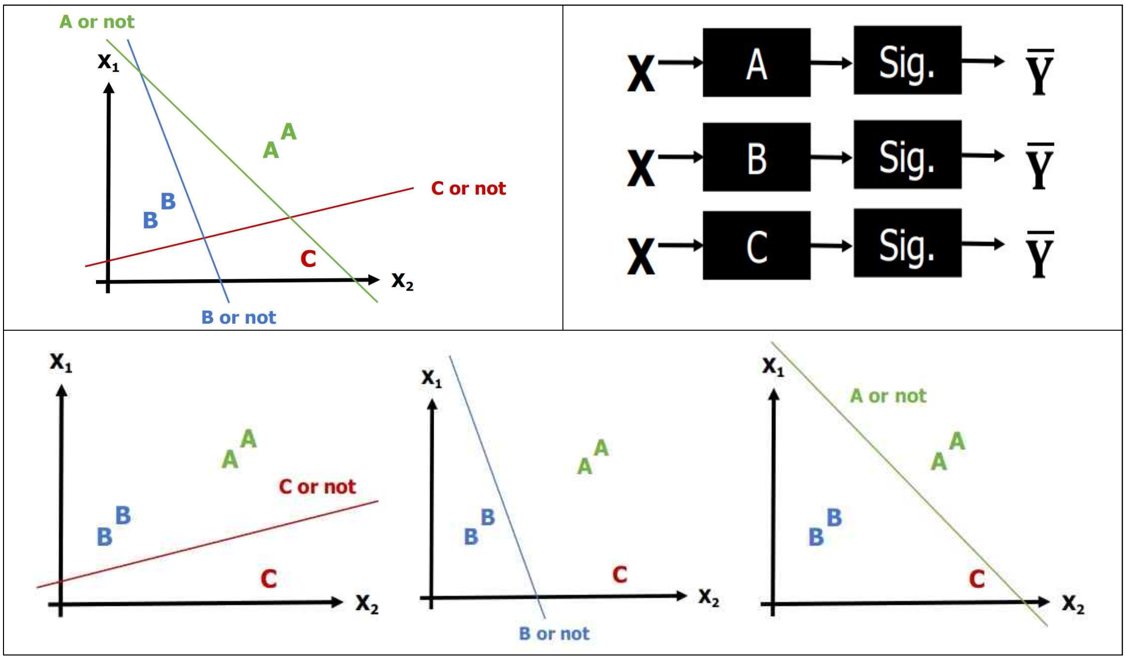
Cost function

: $cost(W) = \frac{1}{m}\sum c(H(X),y) = -\frac{1}{m}\sum\{y\log(H(X)) + (1-y)\log(1-H(X))\}$

Gradient Descent algorithm $W := W - \alpha\frac{\partial}{\partial W}cost(W)$



다중분류(Multi-class Classification)



분류기 A	
$$\begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{A1}x_1 + w_{A2}x_2 + w_{A3}x_3 \end{bmatrix}$$ 분류기 B $$\begin{bmatrix} w_{B1} & w_{B2} & w_{B3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{B1}x_1 + w_{B2}x_2 + w_{B3}x_3 \end{bmatrix}$$ 분류기 C $$\begin{bmatrix} w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{C1}x_1 + w_{C2}x_2 + w_{C3}x_3 \end{bmatrix}$$	$\Rightarrow$ 하나로 합치기 $$\begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{A1}x_1 + w_{A2}x_2 + w_{A3}x_3 \\ w_{B1}x_1 + w_{B2}x_2 + w_{B3}x_3 \\ w_{C1}x_1 + w_{C2}x_2 + w_{C3}x_3 \end{bmatrix}$$

$$H(X) = W^T X = \begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \bar{y}_A \\ \bar{y}_B \\ \bar{y}_C \end{bmatrix}$$

BUT. if 예측값이 $\begin{bmatrix} 2.0 \\ 1.0 \\ 0.1 \end{bmatrix}$ 이라면 sigmoid $\rightarrow$ $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ but 정답은 $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

Sol) 예측값을 [0, 1]로 제한하자

Softmax function: $S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$

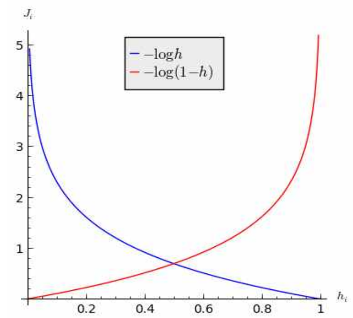
1. 모든 값이 0과 1 사이
2. 전체 합 = 1 (확률 정규화)

$$\mathbf{y} = \begin{bmatrix} 2.0 \\ 1.0 \\ 0.1 \end{bmatrix} \rightarrow \boxed{\text{Softmax}} \xrightarrow{\begin{bmatrix} 0.7 \\ 0.2 \\ 0.1 \end{bmatrix}} \boxed{\text{One-Hot Ecoding}} \rightarrow \overline{\mathbf{Y}} = \begin{bmatrix} 1.0 \\ 0.0 \\ 0.0 \end{bmatrix}$$

Cross-entropy Cost function

예측값 $Softmax(Y) = S(Y) = \overline{Y} = S$

정답 $Ground\ Truth: Y = L$



Cross-entropy

$$D(S, L) = -\sum_i L_i \log(S_i) = -\sum_i L_i \log(\overline{y_i}) = \sum_i L_i * \log(\overline{y_i})$$

$*$ 과 $\odot$: $element-wise$ 곱셈

ex) 정답 $Y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$일 때

1. 참: $\overline{Y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \rightarrow D(S, L) = Y \odot \overline{Y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot -\log \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{bmatrix} \infty \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow 0$

2. 거짓: $\overline{Y} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \rightarrow D(S, L) = Y \odot \overline{Y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot -\log \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{bmatrix} 0 \\ \infty \end{bmatrix} = \begin{bmatrix} 0 \\ \infty \end{bmatrix} \Rightarrow \infty$

정리

Hypothesis $y = H(X) = \frac{1}{1 + e^{-W^T X}}$

Sigmoid function $\sigma(Z) = \frac{1}{1 + e^{-Z}}$

Softmax function $S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$

Cross-entropy

$c(H(X), y) = -y \log(H(x)) - (1 - y) \log(1 - H(x))$

$D(S, L) = -\sum_i L_i \log(S_i)$

Cost function

$\frac{1}{m} \sum c(H(X), y)$

$\frac{1}{m} \sum_i D(S(WX_i + b), L_i)$