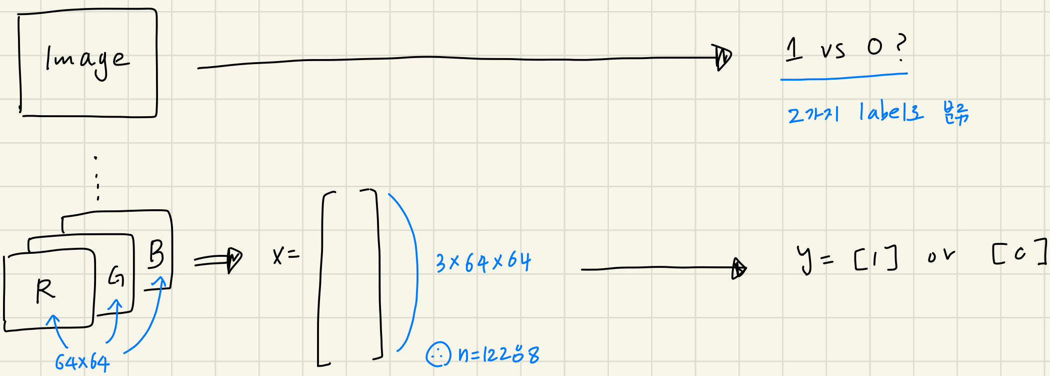
이진 분류



(x, y) , $x \in \mathbb{R}^n$, $y \in \{0, 1\}$

m개의 training sample : $\{ (x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots (x^{(m)}, y^{(m)}) \}$

∴

$$X = \underbrace{\begin{bmatrix} \vdots & \vdots & & \vdots \\ x^{(1)} & x^{(2)} & \cdots & x^{(m)} \\ \vdots & \vdots & & \vdots \end{bmatrix}}_{m} \Bigg\} n \Rightarrow X \in \mathbb{R}^{n \times m}$$

X.shape = $(n \times m)$

$$Y = \underbrace{[\, y^{(1)}, y^{(2)}, \cdots y^{(m)} \,]}_{m} \Big\} 1 \Rightarrow Y \in \mathbb{R}^{1 \times m}$$

Y.shape = $(1 \times m)$

로지스틱 회귀 ← use in Logistic Regression

Given x, we want $\hat{y} = P(1|x)$. x일때 1일 확률

$0 \le \hat{y} \le 1$ ← 조건 충족X

i) $x \in \mathbb{R}^n$

ii) parameters: $w \in \mathbb{R}^n$, $b \in \mathbb{R}$

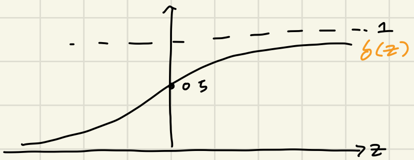
output : $\hat{y} = w^T x + b$ ⇒ 음수일수도 있고, 1보다 큰 수도 있음

∴ $\hat{y} = \underset{\text{시그모이드 함수}}{\sigma}(\underbrace{w^T x + b}_{z})$

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

∴ $z \to \infty$, $\sigma(z) \approx 1$

$z \to -\infty$, $\sigma(z) \approx 0$



⇒ * 선형회귀 *

$\hat{y} = w^T x + b$

* 로지스틱 회귀 *

$\hat{y} = \sigma(w^T x + b)$

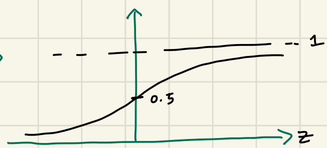
$\hat{y}$ 값이 0~1 범위를 넘어갈 수 있음

로지스틱 회귀의 비용함수

• Logistic Regression

$\hat{y}^{(i)} = \sigma(\overbrace{w^T x^{(i)} + b^{(i)}}^{z})$, $\sigma(z^{(i)}) = \frac{1}{1+e^{-z^{(i)}}}$: $\hat{y}$ 값이 0~1 사이의 값이 되도록 "시그모이드" 함수 적용.

$\sigma(z)$ ⇒



• Loss function ($\mathcal{L}(\hat{y}, y)$)

: $\hat{y}$ (예측값) 과 y (실제값)의 오차의 크기를 나타내는 함수

로지스틱의 오차함수

$\mathcal{L}(\hat{y},y) = -(y\log\hat{y} + (1-y)\log(1-\hat{y}))$

⊖부호 뒤에 있는 값이 커져야 오차함수 전체 값이 작아짐

• If $y=1$, $\mathcal{L}(\hat{y},y) = -\log\hat{y}$ ← we want $\log\hat{y}\uparrow$, ∴ $\hat{y}$ to be large ← 시그모이드 함수에서 가장 큰 값은 1 ⇒ $y \approx \hat{y}$

If $y=0$, $\mathcal{L}(\hat{y},y) = -\log(1-\hat{y})$ ← " $\log(1-\hat{y})\uparrow$, ∴ $\hat{y}$ to be small ← 가장 작은 값은 0 ⇒ $y \approx \hat{y}$

• cost function

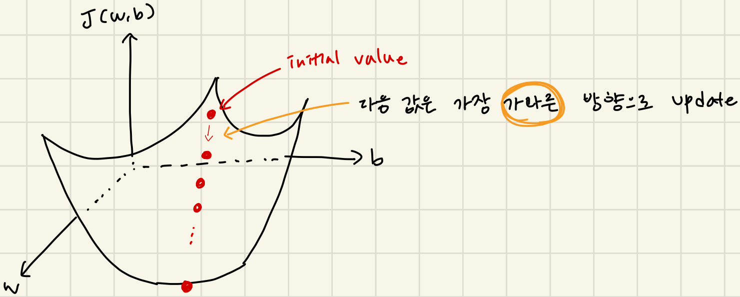
vs Loss function ⇒ 손실함수는 Training set 1개에 대해서 적용

But, cost function : $J(w,b) = \frac{1}{m}\sum^{m} \mathcal{L}(\hat{y},y) = -\frac{1}{m}\sum^{m}[y(\log\hat{y} + (1-y)\log(1-\hat{y})]$

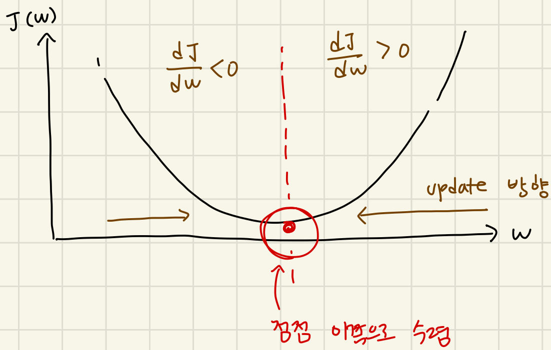
⇒ Training set 모두에 대해서 각각 적용한 값들의 평균

J(w,b)를 minimize 하는 w, b를 찾아야 함

경사하강법



(1) with respect to "w"



Repeat :

update $w = w - \alpha \frac{dJ}{dw}$

learning rate

$\frac{dJ}{dw} > 0$

부호가 ⊖ 가 된다

음의 방향 (왼쪽)으로 w가 update

$\frac{dJ}{dw} < 0$

부호가 ⊕가 된다

양의 방향 (오른쪽)으로 w가 update.

To minimize J(w, b),

we update w, b using Gradient Descent

$$w = w - \alpha \frac{dJ}{dw}$$

$$b = b - \alpha \frac{dJ}{db}$$

계산그래프

Compute $J(a,b,c) = 3(a+bc) = 3(5+3\cdot 2) = 33$

(i) First, compute bc

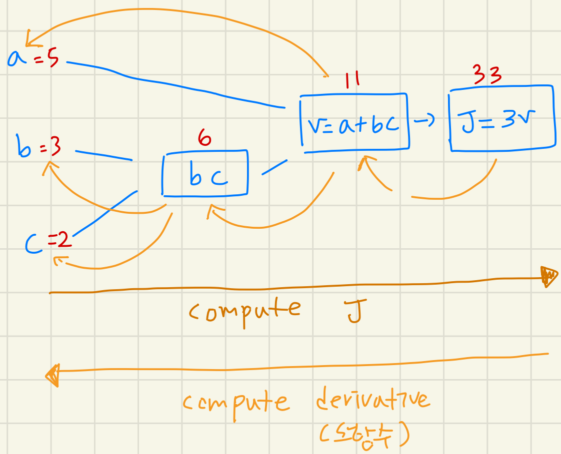
: $u = bc$

(ii) second, compute $a + bc$

: $v = a + u$

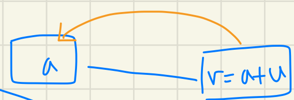
(iii) Finally, compute J.

: $J = 3v$



*compute Derivative

$\frac{dJ}{dv}$ 가 $\frac{dJ}{da}$ 에 영향을 줌

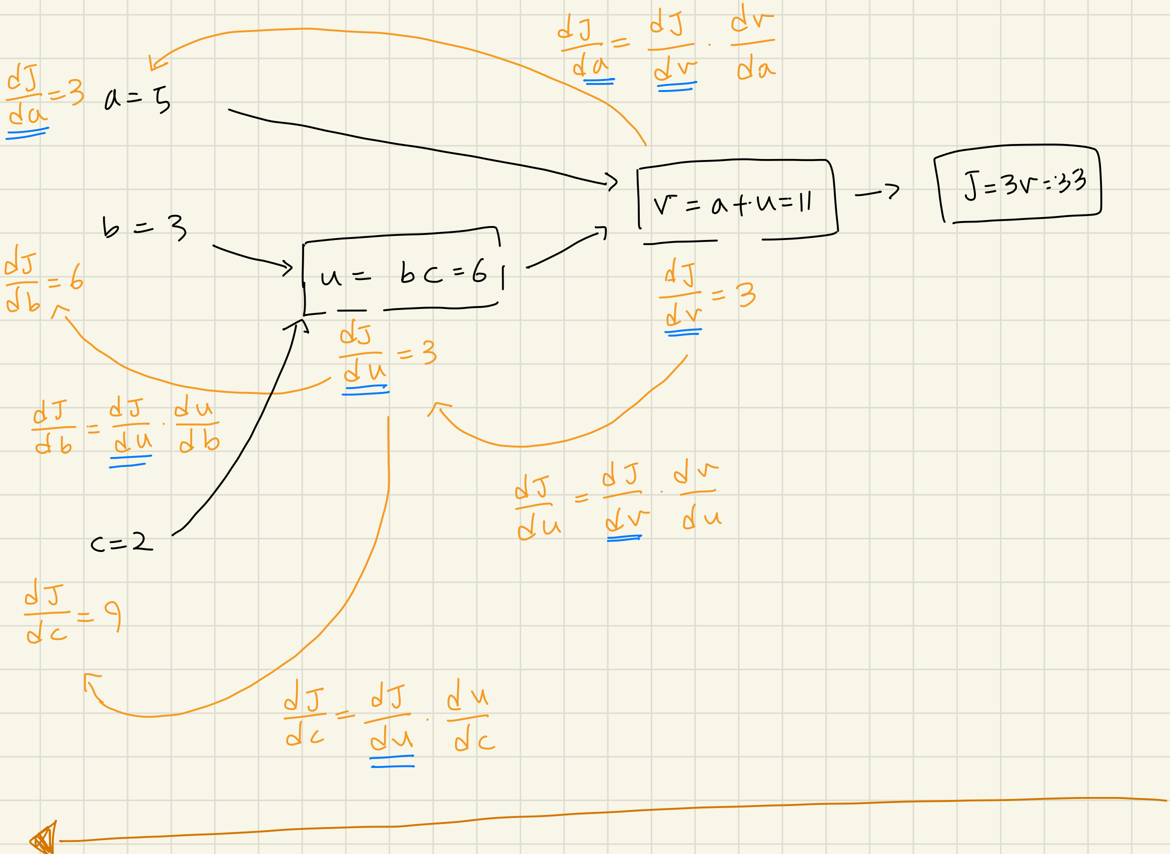


$$\frac{dJ}{dv} = \frac{d(3v)}{dv} = 3$$

$$\frac{dJ}{da} = \frac{d(3v)}{da} = \frac{d(3a+3u)}{da} = 3 = \underbrace{\frac{dJ}{dv}} \cdot \frac{dv}{da}$$

$\because \frac{dv}{da} = 1$

$$\underbrace{\frac{dJ}{dv}}_{3} \cdot \underbrace{\frac{dv}{da}}_{\because 1} = 3$$



도함수 구하는 방향

i.e. $\frac{dJ}{dc}$를 계산하기 위해서는 우선 $\frac{dJ}{du}$를 먼저 계산 함.

로지스틱 회귀의 경사하강법

$z = w^T x + b$

$\hat{y} = a = \sigma(z)$

$\mathcal{L}(a, y) = -(y \log(a) + (1-y)\log(1-a))$

$$[w_1\ w_2]\begin{array}{c} \begin{bmatrix} x_{11}^{(1)} & x_{12}^{(2)} \\ x_{21}^{(1)} & x_{22}^{(2)} \end{bmatrix} \\ \begin{array}{cc} \downarrow & \downarrow \\ x^{(1)} & x^{(2)} \end{array} \end{array} + b$$

x_1

$\frac{d\mathcal{L}}{dw_1} \cdots w_1$

x_2

$\frac{d\mathcal{L}}{dw_2} \cdots w_2$

$\frac{d\mathcal{L}}{db} \cdots b$

$$\boxed{z = w_1x_1 + w_2x_2 + b} \rightarrow \boxed{a = \sigma(z)} \rightarrow \boxed{\mathcal{L}(a, y)}$$

$* (\log_a x)' = \frac{1}{x\ln a}$

minimize $\mathcal{L}(a,y)$

☹ compute 오함수

$\frac{d\mathcal{L}}{da} = \frac{d\mathcal{L}(a,y)}{da}$

$= -\frac{y}{a} + \frac{1-y}{1-a}$

$\frac{d\mathcal{L}}{dz} = \frac{d\mathcal{L}}{da} \cdot \frac{da}{dz}$

$= \frac{d\mathcal{L}}{da} \cdot (a(1-a))$

☹ 역방향 전파의 마지막 단계는

$\mathcal{L}(a,y)$를 minimize 하기 위해 w, b를 얼마나 바꿔야 하는지 계산하는 것

$\Rightarrow \frac{d\mathcal{L}}{dw}$ & $\frac{d\mathcal{L}}{db}$

G·D on m Samples

$J=0,\ dw_1=0,\ dw_2=0,\ db=0$

For $i=1$ to m:

$z^{(i)} = w^T x^{(i)} + b$

$a^{(i)} = \sigma(z^{(i)})$

$J \mathrel{+}= -[y^{(i)}\log(a^{(i)}) + (1-y^{(i)})(\log(1-a^{(i)}))]$

$dz^{(i)} = a^{(i)} - y^{(i)}$

n=2 가정

$dw_1 \mathrel{+}= x_1^{(i)} + dz^{(i)}$

$dw_2 \mathrel{+}= x_2^{(i)} + dz^{(i)}$

$db \mathrel{+}= dz^{(i)}$

$\frac{d\mathcal{L}}{dw}$, $\frac{d\mathcal{L}}{db}$ 를 구하기 위해서 $\frac{d\mathcal{L}}{dz}$를 먼저 계산.

$$[w_1\ w_2]\begin{bmatrix} x_{11}^{(1)} & x_{12}^{(2)} & \cdots & x_{1m}^{(m)} \\ x_{21}^{(1)} & x_{22}^{(2)} & \cdots & x_{2m}^{(m)} \end{bmatrix} + b$$

$w_1 := w_1 - \alpha \underbrace{\frac{dJ}{dw_1}}_{=dw_1}$

$w_2 := w_2 - \alpha \underbrace{\frac{dJ}{dw_2}}_{=dw_2}$

$b := b - \alpha \underbrace{\frac{dJ}{db}}_{=db}$