asgn 2 2023-02-22 박가윤

#1

단어: very, story, disappointing, interesting, bored, fresh

hinge loss + SGD

1. review 1.

$\phi(x) = \{disappointing:1, story:1\}$

$w = (0,0,0,0,0,0)$, $w \cdot \phi(x) = 0$, $w \cdot \phi(x) \cdot y = 0 < 1$

gradient of hinge loss $= -\phi(x)\, y \,[(w \cdot \phi(x))\, y < 1]$

$= -[0,1,1,0,0,0] \cdot (-1) \cdot 1 = [0,1,1,0,0,0]$

$w \leftarrow [0,0,0,0,0,0] - 0.1 \cdot [0,1,1,0,0,0] = [0, -0.1, -0.1, 0, 0, 0]$

2. review 2

$\phi(x) = \{very:1, bored:1\}$

$w = [0, -0.1, -0.1, 0, 0, 0]$, $w \cdot \phi(x) = 0$, $w \cdot \phi(x) \cdot y = 0 < 1$

기울기 $= -[1,0,0,0,1,0] \cdot (-1) \cdot 1 = [1,0,0,0,1,0]$

$w \leftarrow [0, -0.1, -0.1, 0, 0, 0] - 0.1\,[1,0,0,0,1,0] = [-0.1, -0.1, -0.1, 0, -0.1, 0]$

3. review 3

$\phi(x) = \{interesting:1, story:1\} = [0,1,0,1,0,0]$

$w = [-0.1, -0.1, -0.1, 0, -0.1, 0]$

$W \cdot \phi(x) = -0.1 \cdot 1 + 0 \cdot 1 = -0.1$, $W \cdot \phi(x) \cdot y = -0.1 < 1$

기울기 $= -[0,1,0,1,0,0] \cdot (+1) \cdot 1 = [0,-1,0,-1,0,0]$

$W \leftarrow [-0.1, -0.1, -0.1, 0, -0.1, 0] - 0.1\,[0,-1,0,-1,0,0]$

$= [-0.1, 0, -0.1, 0.1, -0.1, 0]$

4. review 4

$\phi(x) = \{$very:1, fresh:1$\} = [1,0,0,0,0,1]$

$W = [-0.1, 0, -0.1, 0.1, -0.1, 0]$, $W \cdot \phi(x) = -0.1$, $W \cdot \phi(x) \cdot y = -0.1 < 1$

기울기: $-[1,0,0,0,0,1] \cdot (+1) \cdot (+1) = [-1,0,0,0,0,-1]$

$W \leftarrow [-0.1, 0, -0.1, 0.1, -0.1, 0] - 0.1[-1,0,0,0,0,-1]$

$= [0, 0, -0.1, 0.1, -0.1, 0.1]$

answer: $W = [0, 0, -0.1, 0.1, -0.1, 0.1]$

#2

1. 2a

Squared Loss $\Rightarrow$ $Loss(x, y, W) = (y - \sigma(W \cdot \phi(x)))^2$

2. 2b

$Loss = (y-p)^2$

① loss를 p에 대해 미분 $\rightarrow$ $\frac{\partial Loss}{\partial p} = (y-p) \cdot 2 \cdot (-1) = -2(y-p)$

② p를 w에 대해 미분

$p = \sigma(z)$, $z = w \cdot \phi(x)$ $\quad \frac{d\sigma(z)}{dz} = \sigma(z)(1-\sigma(z)) = p(1-p)$

$\frac{\partial z}{\partial w} = \phi(x)$, $\frac{\partial p}{\partial w} = p(1-p)\phi(x)$

⇓ chain Rule 사용

$\nabla_w \text{loss}(x, y, w) = -2p(y-p)(1-p)\phi(x)$