1. $w_{k+1} = w_k - \eta \nabla w \, C(w_k)$

$y = f(x) = x^4$

$\nabla w \, C(w_k) = 4x^3$

$x_0 = 2, \quad \nabla w \, C(w_0) = 4 \times 2^3$

$= 32$

$x_1 = x_0 - \eta \nabla w \, C(w_0)$

$= 2 - 0.1 \times 32$

$= -1.2$

$$x = \begin{bmatrix} 1 & 2013 \\ 1 & 2014 \\ 1 & \vdots \\ 1 & \end{bmatrix} \quad (x)$$

2. $x = [1 \quad \text{year}]^T$

$y = f(x, w) = \exp(-x^T w)$ (pred)

$= e^{-x^T w}$

$\nabla w (e^{-x^T w})$

$= -x e^{-x^T w}$

$\nabla (x^T w) = x$

$C(w) = \sum_{i=1}^{m} (f(x, w) - y_i)^2$

sum (pred - y)^2

$$\nabla_w C(w) = \sum_{i=1}^{m} 2\,(f(x,w) - y_i)$$

$$\times \nabla_w f(w_i, w)$$

chain rule

$$= \sum_{i=1}^{m} 2\,(f(x,w) - y_i)\, f(x_i, w)\, x_i$$

sum(2 (pred - y) pred x)

3.

$$y = f(x,w) = x^T w$$

$$\nabla (x^T w) = x$$

$$L(w) = (f(x_i, w) - y_i)^4$$

$$C(w) = \sum_{i=1}^{m} L(w)$$

$$\nabla_w C(w) = \sum_{i=1}^{m} \nabla_w L(w)$$

$$= \sum_{i=1}^{m} \nabla_w (f(x_i, w) - y_i)^4$$

chain rule

$$= \sum_{i=1}^{m} 4\,(f(x_i, w) - y_i)^3\, \nabla_w f(x_i, w)$$

$$= \sum_{i=4}^{m} 4\,(f(x_i, w) - y_i)^3\, x_i$$

4. $f(x, w) = \sigma(x^T w)$

$$\sigma(a) = \frac{1}{1 + \exp(\beta a)}$$

$$a = x^T w$$

$$\nabla_w C(w) = \sum_{i=1}^{m} 4 (f(x_i, w) - y_i)^3$$

$$\times \nabla_w \sigma(x_i^T w)$$ chain rule

$$= \sum_{i=1}^{m} 4 (f(x_i, w) - y_i)^3 \underbrace{\frac{\partial \sigma(a)}{\partial a}}_{?} \underbrace{\nabla_w (x^T w)}_{X}$$

$$\frac{\partial \sigma(a)}{\partial a} = \frac{\partial}{\partial a} \left(\frac{1}{1 + \exp(-\beta a)} \right)$$

$$= - \frac{1}{(1 + \exp(-\beta a))^2} \frac{\partial (1 + \exp(-\beta a))}{\partial a}$$

$$= - \frac{1}{(1 + \exp(-\beta a))^2} \cdot e^{-\beta a} \cdot -\beta$$

$$= \frac{\beta}{(1 + \exp(-\beta a))^2} e^{-\beta a}$$

$$= \frac{\beta}{(1 + e^{-\beta a})^2} \left(e^{-\beta a} + 1 - 1 \right)$$

$$= \beta \, \frac{1 + e^{-\beta a} - 1}{(1 + e^{-\beta a})^2}$$

$\downarrow$ splits

$$= \beta \left(\frac{\cancel{1 +} e^{-\beta a}}{(1 + e^{-\beta a})^{\cancel{2}}} - \frac{1}{(1 + e^{-\beta a})^2} \right)$$

$$= \beta \left(\underbrace{\frac{1}{1 + e^{-\beta a}}}_{\sigma(a)} - \underbrace{\frac{1}{(1 + e^{-\beta a})^2}}_{\sigma^2(a)} \right)$$

$$= \beta \left(\sigma(a) - \sigma^2(a) \right)$$

$$= \beta \, \sigma(a) \left(1 - \sigma(a) \right)$$

$$= \beta \, \sigma(x_i^T w) \left(1 - \sigma(x_i^T w) \right)$$

$$\nabla_w G(w) = \sum_{i=1}^{m} 4\,(f(x_i, w) - y_i)^3$$
$$\times \beta\, \sigma(x_i^T w)\,(1 - \sigma(x_i^T w))\, x_i$$

5.

$$f(x, w) = \sigma(x^T w)$$

$$\sigma(a) = \max(0, a)$$

$$a = x^T w$$

$$\sigma(a) \begin{cases} 0 , & a < 0 \\ a , & a > 0 \end{cases}$$

$$\frac{\partial \sigma(a)}{\partial (a)} \begin{cases} 0 , & a < 0 \\ 1 , & a > 0 \end{cases}$$

$$\frac{\partial \sigma(a)}{\partial (a)} = \delta(a > 0)$$

$$= \delta(x_i^T w > 0)$$

$$\nabla_w (cw) = \sum_{i=1}^{m} 4 (f(x_i, w) - y_i)^3$$

$$\times \delta(x_i^T w > 0) \; x_i$$