①

۹۸۱۰۵۱۳۸ زهرا سوهندل

$\psi(u,v,z) = y$

سؤال ۱)

الف) ماتریس Hessian برای تبدیل ψ را می‌توان به شکل زیر نوشت:

$$H_\psi = \begin{bmatrix} \frac{\partial^2\psi}{\partial u \partial u} & \frac{\partial^2\psi}{\partial u \partial v} & \frac{\partial^2\psi}{\partial u \partial z} \\ \frac{\partial^2\psi}{\partial v \partial u} & \frac{\partial^2\psi}{\partial v \partial v} & \frac{\partial^2\psi}{\partial v \partial z} \\ \frac{\partial^2\psi}{\partial z \partial u} & \frac{\partial^2\psi}{\partial z \partial v} & \frac{\partial^2\psi}{\partial z \partial z} \end{bmatrix} \quad ①$$

برای گرادیان این تبدیل داریم:

$$\nabla\psi = \begin{bmatrix} \frac{\partial\psi}{\partial u} \\ \frac{\partial\psi}{\partial v} \\ \frac{\partial\psi}{\partial z} \end{bmatrix} = \begin{bmatrix} \psi_1(u,v,z) \\ \psi_2(u,v,z) \\ \psi_3(u,v,z) \end{bmatrix}$$

برای ژاکوبین $\nabla\psi$ داریم:

$$J(\nabla\psi) = \begin{bmatrix} \nabla^T\psi_1 \\ \nabla^T\psi_2 \\ \nabla^T\psi_3 \end{bmatrix} = \begin{bmatrix} \frac{\partial^2\psi}{\partial u \partial u} & \frac{\partial^2\psi}{\partial u \partial v} & \frac{\partial^2\psi}{\partial u \partial z} \\ \frac{\partial^2\psi}{\partial v \partial u} & \frac{\partial^2\psi}{\partial v \partial v} & \frac{\partial^2\psi}{\partial v \partial z} \\ \frac{\partial^2\psi}{\partial z \partial u} & \frac{\partial^2\psi}{\partial z \partial v} & \frac{\partial^2\psi}{\partial z \partial z} \end{bmatrix} \quad ②$$

در نتیجه داریم با توجه به $② = ①,②$

$$H_\psi = J(\nabla\psi) \quad \boxed{\checkmark}$$

ب) ؟$\left(\frac{\partial y}{\partial x} = A\right)$ (ب)

مقدار درایه ij در ماتریس $\frac{\partial y}{\partial x}$ را محاسبه می‌کنیم

$$\left(\frac{\partial y}{\partial x}\right)_{ij} = \frac{\partial y_i}{\partial x_j} = \frac{\partial\left(\sum_{k=1}^{n} a_{ik} x_k\right)}{\partial x_j} = a_{ij}$$

طبق فرم نشان داده

$$\Rightarrow \left(\frac{\partial y}{\partial x}\right)_{ij} = a_{ij} \Rightarrow \frac{\partial y}{\partial x} = A \quad \boxed{\checkmark}$$

(22

$$\frac{\partial y}{\partial z} = \begin{bmatrix} \frac{\partial y_1}{\partial z} \\ \vdots \\ \frac{\partial y_m}{\partial z} \end{bmatrix} = \begin{bmatrix} \frac{\partial(\langle A_1 , x \rangle)}{\partial z} \\ \vdots \\ \frac{\partial(\langle A_m , x \rangle)}{\partial z} \end{bmatrix} = \begin{bmatrix} \frac{\partial\left(\sum_{k=1}^{n} a_{1k} x_k\right)}{\partial z} \\ \vdots \\ \frac{\partial\left(\sum_{k=1}^{n} (a_{mk} \cdot x_k)\right)}{\partial z} \end{bmatrix}$$

با توجه به اینکه داریم: $\frac{\partial x}{\partial z} = \left[\frac{\partial x_1}{\partial z} \cdots \frac{\partial x_n}{\partial z}\right]^T$

$$= \begin{bmatrix} \sum_{k=1}^{n} a_{1k} \frac{\partial x_k}{\partial z} \\ \vdots \\ \sum_{k=1}^{n} a_{mk} \frac{\partial x_k}{\partial z} \end{bmatrix} = \begin{bmatrix} \langle A_1 , \frac{\partial x}{\partial z} \rangle \\ \vdots \\ \langle A_m , \frac{\partial x}{\partial z} \rangle \end{bmatrix} = \begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix} \left(\frac{\partial x}{\partial z}\right) = A \cdot \frac{\partial x}{\partial z}$$

$$\Rightarrow \frac{\partial y}{\partial z} = A \cdot \frac{\partial x}{\partial z} \quad \boxed{\checkmark}$$

$$\alpha = y^T A x = [y_1 \cdots y_m]\begin{bmatrix} r_1 \\ \vdots \\ r_m \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{m} y_i \langle r_i . x \rangle \quad \text{(iii}$$

$$\boxed{\frac{\partial \alpha}{\partial x} = \left[\frac{\partial \alpha}{\partial x_1} \cdots \frac{\partial \alpha}{\partial x_n}\right]} \qquad = \sum_{i=1}^{m} y_i \left(\sum_{k=1}^{n} a_{ik} x_k\right) = M$$

$$\frac{\partial \alpha}{\partial x_j} = \frac{\partial M}{\partial x_j} = \sum_{i=1}^{m} y_i (a_{ij}) \quad (1)$$

از طرفی داریم: $(y^T A)_j = \left([y_1 \cdots y_m]\begin{bmatrix} | & & | & & | \\ c_1 & \cdots & c_k & \cdots & c_n \\ | & & | & & | \end{bmatrix}\right)_j = \langle y . c_j \rangle$

$$= \sum_{i=1}^{m} y_i a_{ij} \quad (2)$$

$$(1),(2) \Rightarrow \forall j ; \frac{\partial \alpha}{\partial x_j} = (y^T A)_j \Rightarrow \frac{\partial \alpha}{\partial x} = y^T A \quad \boxed{\checkmark}$$

$$\frac{\partial \alpha}{\partial y} = \left[\frac{\partial \alpha}{\partial y_1} \cdots \frac{\partial \alpha}{\partial y_m}\right]$$

$$\frac{\partial \alpha}{\partial y_j} = \frac{\partial M}{\partial y_j} = \sum_{k=1}^{n} a_{jk} x_k \quad (1)$$

از طرفی داریم: $(x^T A^T)_j = \left([x_1 \cdots x_n]\begin{bmatrix} | & & | \\ r_1 & \cdots & r_n \\ | & & | \end{bmatrix}\right)_j = \langle x . r_j \rangle$

$$= \sum_{k=1}^{n} x_k a_{jk} \quad (2)$$

$$(1),(2) \Rightarrow \frac{\partial \alpha}{\partial y_j} = (x^T A^T)_j \Rightarrow \frac{\partial \alpha}{\partial y} = x^T A^T \quad \boxed{\checkmark}$$

(iv

$$\alpha = y^T x = \sum_{i=1}^{n} y_i x_i$$

$$\frac{\partial \alpha}{\partial z} = \frac{\partial \sum_{i=1}^{n} y_i x_i}{\partial z} = \sum_{i=1}^{n} \frac{\partial (x_i y_i)}{\partial z} = \sum_{i=1}^{n} \frac{\partial x_i}{\partial z} y_i$$

$$+ \sum_{i=1}^{n} x_i \frac{\partial y_i}{\partial z} \rightarrow \frac{\partial y}{\partial z} = \begin{bmatrix} \frac{\partial y_1}{\partial z} \\ \vdots \\ \frac{\partial y_n}{\partial z} \end{bmatrix} \quad \frac{\partial x}{\partial z} = \begin{bmatrix} \partial x_1 / \partial z \\ \vdots \\ \partial x_n / \partial z \end{bmatrix}$$

$$\Rightarrow \frac{\partial \alpha}{\partial z} = \langle y \cdot \frac{\partial x}{\partial z} \rangle + \langle x \cdot \frac{\partial y}{\partial z} \rangle$$

$$\Rightarrow \frac{\partial \alpha}{\partial z} = y^T \frac{\partial x}{\partial z} + x^T \frac{\partial y}{\partial z} \quad \boxed{\checkmark}$$

(v

با توجه به تعریف وارون پذیری $\Rightarrow \quad A^{-1} A = I$

مشتق میگیریم $\Rightarrow \quad A^{-1} \frac{\partial A}{\partial \alpha} + \frac{\partial A^{-1}}{\partial \alpha} A = \frac{\partial I}{\partial \alpha} = 0$

$$\Rightarrow \frac{\partial A^{-1}}{\partial \alpha} A = - A^{-1} \frac{\partial A}{\partial \alpha}$$

$$\xRightarrow{\times A^{-1}} \frac{\partial A^{-1}}{\partial \alpha} = - A^{-1} \frac{\partial A}{\partial \alpha} A^{-1} \quad \boxed{\checkmark}$$

ج) ابتدا برای اینکه یک مقدار ویژه باشد باید داشته باشیم:

m: is an eigenvalue $\Rightarrow \exists v \neq 0 \; ; \; Av = mv$

حال کافی است نشان دهیم که وجود دارد بردار v طوری که $Av = mv$

بردار v را به صورت $v^T = [1 1 1 \dots 1]_n$ در نظر بگیرید.

حال داریم:

$$Av = \begin{bmatrix} - r_1 - \\ \vdots \\ - r_n - \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} a_{1j} \\ \vdots \\ \sum_{j=1}^{n} a_{nj} \end{bmatrix} = \begin{bmatrix} m \\ \vdots \\ m \end{bmatrix} = mv$$

$\Rightarrow Av = mv \Rightarrow$ m یک مقدار ویژه برای ماتریس A می‌باشد.

د) برای اینکه این ترم جریمه $R(w) = w^T S w$ از نامتقارن شدن وزن ها جلوگیری کند در نتیجه باید داشته باشیم
مینیمم شدن $R(w)$ مشروط با متقارن شدن w شود. یعنی در مقادیر مینیمم برای $R(w)$ داشته باشیم
$w_1 = w_2$ باشد. به صورت زیر عمل می‌کنیم که در مشتق برابر صفر برای $R(w)$ شرط $w_1 = w_2$ را داشته باشیم.
S را طوری تعیین می‌کنیم که این اتفاق بیافتد.

$$w^T S w = [w_1 \quad w_2]\begin{bmatrix} S_1 & S_2 \\ S_3 & S_4 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = [w_1 S_1 + w_2 S_3 \quad w_1 S_2 + w_2 S_4]\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$\Rightarrow \quad R(w) = w_1^2 S_1 + w_1 w_2 (S_3 + S_2) + w_2^2 S_4$$

$$\frac{\partial R}{\partial w_1} = 0 \Rightarrow 2 w_1 S_1 + w_2 (S_3 + S_2) = 0 \Rightarrow \frac{w_1}{w_2} = \frac{-(S_3 + S_2)}{2 S_1}$$

$$\frac{\partial R}{\partial w_2} = 0 \Rightarrow w_1 (S_3 + S_2) + 2 w_2 S_4 = 0 \Rightarrow \frac{w_1}{w_2} = \frac{-(S_2 + S_3)}{2 S_4}$$

اگر $w_1 = w_2 \Rightarrow \begin{cases} S_1 = \dfrac{-S_2 - S_3}{2} \\ , \; S_4 = \dfrac{-S_2 - S_3}{2} \end{cases}$ $\Rightarrow$ برای مثال $\Rightarrow S = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$

سوال ۲) برای نقاط استاتیک تابع را به دست آوردن ابتدا مقدار گرادیان تابع f مشخص می‌کنیم.

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 6x_1 - 3x_2 + 12x_1^2 + 4x_1^3 \\ 4x_2 - 3x_1 \end{bmatrix} = \begin{bmatrix} f_{x_1} \\ f_{x_2} \end{bmatrix}$$

$$\nabla f = 0 \Rightarrow \left\{ \begin{array}{l} 6x_1 - 3x_2 + 12x_1^2 + 4x_1^3 = 0 \\ 4x_2 - 3x_1 = 0 \Rightarrow x_2 = \frac{3x_1}{4} \end{array} \right\} \Rightarrow 6x_1 - \frac{9}{4}x_1 + 12x_1^2 + 4x_1^3 = 0$$

$$\Rightarrow x_1 (4x_1^2 + 12x_1 + 3.75) = 0 \Rightarrow x_1 = 0 \text{ یا } x_1 = \frac{-12 \pm \sqrt{84}}{8}$$

$$\Rightarrow \left\{ \begin{array}{l} x_1 = 0 \\ x_1 = \frac{-6 + \sqrt{21}}{4} \\ x_1 = \frac{-6 - \sqrt{21}}{4} \end{array} \right. \text{ و } x_2 = \frac{3}{4} x_1$$

نقاط استاتیک تابع f نقاط بالا هستند $\Uparrow$.

حال ۳ حالت مختلف را بررسی می‌کنیم که هر نقطه چه نوع نقطه استاتیکی می‌باشد.

ابتدا مقدار D را به صورت پارامتریک می‌نویسیم.

$$D = f_{x_1x_1} \cdot f_{x_2x_2} - f_{x_1x_2}^2$$

$$\left. \begin{array}{l} f_{x_1x_1} = \frac{\partial^2 f}{\partial x_1 \partial x_1} = 6 + 24x_1 + 12x_1^2 \\ f_{x_2x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_2} = 4 \\ f_{x_1x_2} = -3 \end{array} \right\} \Rightarrow \begin{array}{l} D = 24 + 96x_1 + 48x_1^2 - 9 \\ D = 48x_1^2 + 96x_1 + 15 \end{array}$$

$$D(0) = 15 > 0 \quad \text{و} \quad \frac{\partial^2 f}{\partial x_1^2}(0) = 6 > 0 \Rightarrow x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
نقطه استاتیک کمینه محلی است.

$$D\left(\frac{\sqrt{21}-6}{4}\right) < 0 \Rightarrow x = \begin{bmatrix} x_1 \\ 0.75x_1 \end{bmatrix} = \begin{bmatrix} (\sqrt{21}-6)/4 \\ \frac{3}{16}(\sqrt{21}-6) \end{bmatrix}$$
یک نقطه استاتیک نقطه زینی می‌باشد.

$$D\left(\frac{-\sqrt{21}-6}{4}\right) > 0 \quad \text{و} \quad \frac{\partial^2 f}{\partial x_1^2} > 0 \Rightarrow x = \begin{bmatrix} x_1 \\ 0.75x_1 \end{bmatrix} = \begin{bmatrix} (-\sqrt{21}-6)/4 \\ 3(-\sqrt{21}-6)/16 \end{bmatrix}$$
نقطه استاتیک مقابل یک نقطه کمینه محلی است.

سوال ۲)

بهینه سازی در توابع غیر محدب:

الف)

$$f(x) = 2x_1^2 + 2x_2^2 - 17x_2 \cos(0.2\pi x_1) - x_1 x_2$$

$$f' = \nabla(f) = \begin{bmatrix} 4x_1 + 17x_2(0.2\pi)\sin(0.2\pi x_1) - x_2 \\ 4x_2 - 17\cos(0.2\pi x_1) - x_1 \end{bmatrix}$$

$$f'' = H(f) = \begin{bmatrix} 4 + 17(0.2\pi)^2 x_2 \cos(0.2\pi x_1) & 17(0.2\pi)\sin(0.2\pi x_1) - 1 \\ 17(0.2\pi)\sin(0.2\pi x_1) - 1 & 4 \end{bmatrix}$$

در نقطه شروع $(0,0)$ داریم:

$$x_{n+1} = x_n - \alpha H^{-1}(x_n) f'(x_n)$$

$$x_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow f'(x_0) = \nabla f(x_0) = \begin{bmatrix} 0 \\ -17 \end{bmatrix}$$

$$\Rightarrow f''(x_0) = H_f(x_0) = \begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix}$$

$$H_f^{-1}(x_0) = \frac{1}{15}\begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$$

$$x^{(1)} = x_0 - \alpha H^{-1}(x_0) f'(x_0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \alpha\left(\frac{1}{15}\begin{bmatrix} 4 & 1 \\ 4 & 4 \end{bmatrix}\begin{bmatrix} 0 \\ -17 \end{bmatrix}\right)$$

$$x^{(1)} = x^{0} - \alpha \begin{bmatrix} -1.13 \\ -4.53 \end{bmatrix}$$

$$x^{(1)} = \begin{bmatrix} 0 + 1.13\alpha \\ 0 + 4.53\alpha \end{bmatrix} = \begin{bmatrix} 1.13\alpha \\ 4.53\alpha \end{bmatrix}$$ مقدار جدید نقطه شروع پس از یکبار آپدیت کردن.

ج)

خط $\boxed{1}$ = در این خط گرادیان تابع هدف را نسبت به پارامترهای θ_{t-1} حساب می‌کنیم و g_t را برابر آن قرار می‌دهیم.

خط $\boxed{2}$ = مومنتوم را در زمان t، m_t حساب می‌کنیم به این شکل که به صورت وزن‌دار با وزن‌های β_1 و $(1-\beta_1)$ به ترتیب مقدار مومنتوم قبلی و گرادیان جدید را با هم جمع کرده تا مومنتوم در لحظه زمان t را حساب کنیم. این خط مقدار m_t baised را محاسبه کردیم.

خط $\boxed{3}$ = در این بخش v_t را حساب می‌کنیم به این صورت که مقدار v_{t-1} را با ضریب β_2 با مقدار g_t^2 با ضریب $(1-\beta_2)$ جمع می‌کنیم و مقدار v_t را حساب می‌کنیم (آپدیت می‌کنیم).
$\hookrightarrow$ second momentum گرادیان پارامترهاست.

خط $\boxed{4}$ = در این خط مقدار m_t unbaised یعنی $\hat{m}_t$ را حساب می‌کنیم.

خط $\boxed{5}$ = با ضرب $\frac{1}{1-\beta_2^t}$ در مقدار v_t، مقدار بدون بایاس $\hat{v}_t$ را محاسبه می‌کنیم.

خط $\boxed{6}$ = در این خط مقدار θ را آپدیت می‌کنیم به این صورت که به جای استفاده مستقیم از گرادیان از $\hat{m}_t$ ، $\hat{v}_t$ که مقادیر بدون بایاس همراه با مومنتوم هستند استفاده می‌کنیم.
(α همون learning rate میباشد.)

دلیل اینکه چرا m_t به سمت صفر بایاس دارد؟

زیرا $m_0 = 0$ می‌باشد که β_1 معمولاً عددی نزدیک به 1 انتخاب می‌شود که در نتیجه $m_1 = (1-\beta_1) g_t$ که $(1-\beta_1)$ مقدار کمی نزدیک به صفر دارد. که هم $(1-\beta_1)$ و هم g_t مقادیر کوچک نزدیک صفر دارند که در نتیجه در استپ‌های بعدی نیز به سمت صفر میل می‌کنند.

چرا در $\hat{m}_t$ به این شکل می‌خوریم؟

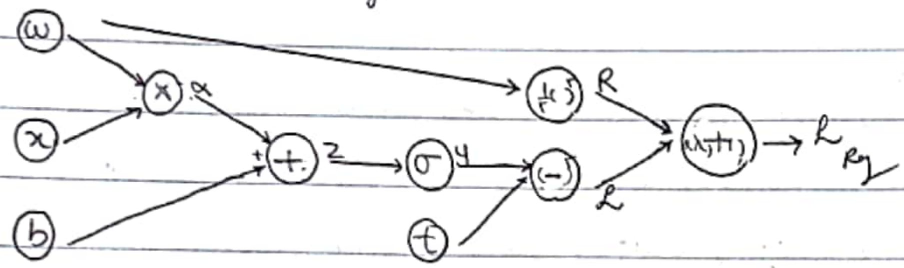
زیرا ضرب کردن $\frac{1}{1-\beta_1^t}$ تاثیری که هم $(1-\beta_1)$ باعث میل به سمت صفر می‌شد را از بین می‌برد و مقدار بزرگی را در m_t ضرب می‌کند و باعث می‌شود دیگر به سمت صفر بایاس نشویم.

باعث می‌شد مقدار مومنتوم بولد و هم مقدار m_t در شروع مقادیر بزرگ بگیرند والبته اونقدر هم بزرگ نه! طوری نیست که مشکل آموزش‌های به در مراحل ابتدایی اولیه به مشکل بخورد.

سوال ۳) الف)

$z = \underbrace{wx}_{\alpha} + b \qquad y = \sigma(z) \qquad \mathcal{L} = \frac{1}{2}(y-t)^2 \qquad R = \frac{1}{2}w^2$

$\mathcal{L}_{reg} = \mathcal{L} + \lambda R$



$\frac{\partial \mathcal{L}_{reg}}{\partial R} = \lambda \qquad \frac{\partial \mathcal{L}_{reg}}{\partial y} = \frac{\partial \mathcal{L}_{reg}}{\partial \mathcal{L}} \cdot \frac{\partial \mathcal{L}}{\partial y} = 1 \times (y-t)$

$\frac{\partial \mathcal{L}_{reg}}{\partial \mathcal{L}} = 1 \qquad \frac{\partial \mathcal{L}_{reg}}{\partial t} = \frac{\partial \mathcal{L}_{reg}}{\partial \mathcal{L}} \cdot \frac{\partial \mathcal{L}}{\partial t} = 1 \times (-(y-t))$

$\frac{\partial \mathcal{L}_{reg}}{\partial z} = \frac{\partial \mathcal{L}_{reg}}{\partial y} \times \frac{\partial y}{\partial z} = (y-t) \cdot (\sigma(z))(1-\sigma(z))$

$\frac{\partial \mathcal{L}_{reg}}{\partial \alpha} = \frac{\partial \mathcal{L}_{reg}}{\partial z} \cdot \underbrace{\frac{\partial z}{\partial \alpha}}_{1} = (y-t)\,\sigma(z)\,(1-\sigma(z)) \times 1$

$\frac{\partial \mathcal{L}_{reg}}{\partial b} = \frac{\partial \mathcal{L}_{reg}}{\partial z} \cdot \frac{\partial z}{\partial b} = (y-t)\,\sigma(z)\,(1-\sigma(z)) \times 1$

$\frac{\partial \mathcal{L}_{reg}}{\partial x} = \frac{\partial \mathcal{L}_{reg}}{\partial \alpha} \cdot \underbrace{\frac{\partial \alpha}{\partial x}}_{w} = (y-t)\,\sigma(z)\,(1-\sigma(z))\, w$

$\frac{\partial \mathcal{L}_{reg}}{\partial w} = \frac{\partial \mathcal{L}_{reg}}{\partial \alpha} \frac{\partial \alpha}{\partial w} + \frac{\partial \mathcal{L}_{reg}}{\partial R} \cdot \frac{\partial R}{\partial w}$

$= (y-t)\,\sigma(z)(1-\sigma(z)) \cdot x + \lambda \cdot w$

ب) علت رندوم مقداردهی دادن پارامترها به دلیل استفاده از SGD درفرآیند آموزش مدل می باشد. که نیاز به رندوم بودن پارامترها می باشد تا به نقطه کمینه کلی برسیم.

اگر پارامترها مقدار ثابت دهیم در ابتدا، باعث می‌شود پیچیدگی مدل کم شود و فقط یک چیز را آموزش ببیند.

و همچنین رندوم مقداردهی کردن باعث می‌شود تقارن شبکه را بشکند و اجازه دهد شبکه فیچرهای مختلفتری را آموزش ببیند و همچنین باعث افزایش دقت مدل می‌شود.

پس استفاده از مقادیر کوچک به دلیل اینکه در شبکه اگر از مقادیر بزرگ استفاده کنیم چون در فرآیند back propagation از ضرب مشتقات استفاده می‌شود باعث gradiant exploding می‌شود.

همچنین چون از تابع فعال‌ساز مانند σ استفاده می‌کنیم، در مقادیر بزرگ سو شیب نزدیک به صفر دارد. و مشتقات را صفر می‌کند و باعث می‌شود شبکه learn نشود، البته در شبکه‌های عمیق‌تر ممکن است باعث vanishing gradiant شود استفاده از مقادیر خیلی کوچک به عنوان مقدار اولیه برای وزن‌ها.

ج.

$x=1 \quad , \quad t=2 \quad , \quad w=0.5 \quad , \quad b=1 \quad , \quad \lambda=1$

$\alpha=0.1 \leftarrow$ learning rate

البته که λ یک نوع هایپرپارامتر باشد. که با دیتای validation باید تنظیم شود.

$$\mathcal{L}_{Reg} = \mathcal{L} + \lambda R$$

$$= \frac{1}{2}(y-t)^2 + \lambda \frac{1}{2} w^2 \qquad a = xw = 0.5,\ \underbrace{xw+b}_{z=} = 1.5 \qquad y-t = -1.82$$

$$= \frac{1}{2}\left[\left(\sigma(-(\underbrace{wx+b}_{1.5})-2)\right)^2 + \lambda w^2\right] \qquad y = \sigma(z) = 0.817$$

$$= \frac{1}{2}\left[\sigma(1.5)^2 + 0.25\right] = 0.834$$

$$w' = w + \alpha \frac{\partial \mathcal{L}_{Reg}}{\partial w}$$

$$b' = b + \alpha \frac{\partial \mathcal{L}_{Reg}}{\partial b}$$

$$w' = 0.5 + 0.1\left[(y-t)\sigma(z)(1-\sigma(z))x + \lambda w\right]$$

$$= 0.5 + 0.1\left[(-1.82)(0.817)(0.183)1 + 1 \times 0.5\right]$$

$$= 0.532$$

$$b' = 1 + 0.1\left[(y-t)\sigma(z)(1-\sigma(z))\right]$$

$$= 1 + 0.1\left[(-1.82)(0.817)(0.183)\right]$$

$$= 0.982$$

$$\Rightarrow w' = 0.532 \quad , \quad b' = 0.983$$

$$(w,b) = (0.5, 1)$$