$$\nabla_{w_0} L = \frac{\partial L}{\partial w_0} = \frac{\partial L}{\partial f} \times \frac{\partial f}{\partial q} \times \frac{\partial q}{\partial r} \times \frac{\partial r}{\partial s_1} \times \frac{\partial s_1}{\partial w_0}$$
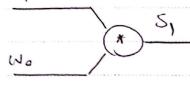
binary cross entropy

$$\frac{\partial L}{\partial f} = \frac{\partial \left(-y \log f - (1-y) \log(1-f)\right)}{\partial f} = \left(\frac{-y}{f}\right) + \left(\frac{1-y}{1-f}\right)$$

$$\frac{\partial f}{\partial q} = (1-f)(f)$$

$$\frac{\partial q}{\partial r} = 1 \quad , \quad \frac{\partial r}{\partial s_1} = 1 \quad , \quad \frac{\partial s_1}{\partial w_0} = n_1$$

$$\Rightarrow \frac{\partial L}{\partial w_0} = \left[\frac{-y}{f} + \frac{1-y}{1-f}\right] f(1-f) \times n_1 = (f-y) n_1$$

SGD , learning_rate $= 0/01$, نمونه F: $(1,٢),(٣,٤),(1,٤),(٢,٣)$

$\downarrow$

$n_{t+1} = n_t - \alpha \nabla f(n_t)$

$w_0 = 1 \times 10^{-5}, \quad w_1 = -٢ \times 10^{-5}, \quad w_٢ = ٣ \times 10^{-5}$

$$f(W, \underbrace{n_1 = 121, n_٢ = 14/8}_{n=1}) = \frac{1}{1+e^{-(10^{-5}\times 121 - ٢\times 10^{-5}\times 14/8 + ٣\times 10^{-9})}} = 0/500$$

$$f(W, \underbrace{n_1 = 114, n_٢ = 15/٢}_{n=٢}) = \frac{1}{1+e^{-(10^{-5}\times 114 - ٢\times 10^{-5}\times 15/٢ + ٣\times 10^{-5})}} = 0/500$$

$$f(W, \underbrace{n_1 = ٢10, n_٢ = 9/4}_{n=٣}) = \frac{1}{1+e^{-(10^{-5}\times ٢10 - ٢\times 10^{-5}\times 9/4 + ٣\times 10^{-5})}} = 0/5..$$

$$f(W, \underbrace{n_1 = 195, n_٢ = 8/1}_{n=4}) = \frac{1}{1+e^{-(10^{-5}\times 195 - ٢\times 10^{-5}\times 8/1 + ٣\times 10^{-9})}} = 0/500$$

خروجی داده ها قبل از آپیت کنیم.

① گام

$\nabla_{w_0} L(f_1, y_1) = (f_1 - y_1)\times n_1 = (0/500 - 0)\times 121 = 40/5$

$\nabla_{w_0} L(f_٢, y_٢) = (f_٢ - y_٢)\times n_1 = (0/500 - 0)\times 114 = 57$

$\nabla_{w_0} L = \frac{1}{٢}(40/5 + 57) = 58/75$

$\Rightarrow w_0 = w_0 - \alpha \nabla_{w_0} L = 10^{-5} - (0/01)(58/75) = -0/587$

برای $\Delta w_1\ell$ که همان استفاده رابطه‌ی قبلی داریم : $\Delta w_1\ell = (f-y)u_2$

$, \Delta w_2\ell = \dfrac{\partial \ell}{\partial f} \times \dfrac{\partial f}{\partial q} \times \dfrac{\partial q}{\partial w_2} = \left[\dfrac{-y}{f} + \dfrac{1-y}{1-f}\right] f(1-f) \times 1$

$= (f-y)$

$\Delta w_1 L(f_1, y_1) = (f_1 - y_1) \times u_2 = (0/5 - 0) \times 16/8 = 8/4$

$\Delta w_1 L(f_2, y_2) = (f_2 - y_2) u_2 = (0/5 - 0) \times 15/2 = 7/4$

$\Delta w_1 L = \dfrac{1}{2}(8/4 + 7/4) = 8$

$\Rightarrow w_1 = w_1 - \alpha \Delta w_1 L = -2 \times 10^{-5} - (0/01)(8) = -0/08.$

$\Delta w_2\ell(f_1, y_1) = (f_1 - y_1) = 0/5$

$\Delta w_2 L(f_2, y_2) = (f_2 - y_2) = 0/5$

$\Delta w_2 L = \dfrac{1}{2}(0/5 + 0/5) = 0/5$

$\Rightarrow w_2 = w_2 - \alpha \Delta w_2 \ell = 2 \times 10^{-5} - (0/01)(0/5) = -0/004$

$$\Rightarrow \omega_0 = -0/۵۸۷, \quad \omega_1 = -0/۰۸, \quad \omega_۲ = -0/۰۰۴$$

$$f_۱ = \cfrac{1}{1+e^{-(-0/۵۸۷ \times ۱۲۱ - 0/۰۸ \times ۱۴/۸ - 0/۰۰۴)}} = ۳/۴۹ \times ۱۰^{-۳۴}$$

$$f_۲ = \cfrac{1}{1+e^{-(-0/۵۸۷ \times ۱۱۴ - 0/۰۸ \times ۱۵/۲ - 0/۰۰۴)}} = ۲/۵۵ \times ۱۰^{-۳۰}$$

$$f_۳ = \cfrac{1}{1+e^{-(-0/۵۸۷ \times ۲۱۰ - 0/۰۸ \times ۹/۴ - 0/۰۰۴)}} = ۱/۳۴ \times ۱۰^{-۵۴}$$

$$f_۴ = \cfrac{1}{1+e^{-(-0/۵۸۷ \times ۱۹۵ - 0/۰۸ \times ۸/۱ - 0/۰۰۴)}} = ۱/۰۱ \times ۱۰^{-۵۰}$$

با توجه به اینکه داده هایی که در این تمرین به الگوریتم، بهینه سازمان نشان دادهایم

داده های ۱ و ۲ هستند و برحسب آن ها هر دو ۰ می باشد آیریت سعی

به گونه ای انجام می شود که خروجی سگموید برای در داده ۱ ، ۲ به ۰ نزدیک

شود. اما این آیریت در پارامترها منجر به خروجی گرفتگ برای داده های

۳ و ۴ می شود بحالیم برحسب این در داده ۱ می باشد.

سوال ۳

$$\nabla_{w_0} L(f_{\gamma}, y_{\gamma}) = (f_{\gamma} - y_{\gamma}) u_1 = (1/\gamma 4 \times 1 \cdot^{-\delta\gamma} - 1) \times \gamma 1 \cdot \simeq -\gamma 1 \cdot$$

$$\nabla_{w_0} L(f_{\gamma}, y_{\gamma}) = (f_{\gamma} - y_{\gamma}) u_1 = (1/\cdot 1 \times 1 \cdot^{-\delta\cdot} - 1) \times 19\delta \simeq -19\delta$$

$$\nabla_{w_0} L = \frac{1}{\gamma}(-\gamma 1 \cdot - 19\delta) = -\gamma \cdot \gamma/\delta$$

$$w_0 = -\cdot/\delta \Lambda V - (\cdot/\cdot 1)(-\gamma \cdot \gamma/\delta) = 1/\gamma \gamma \Lambda$$

$$\nabla_{w_1} L(f_{\gamma}, y_{\gamma}) = (f_{\gamma} - y_{\gamma}) u_\gamma = (1/\gamma 4 \times 1 \cdot^{-\delta\gamma} - 1) \times 9/\gamma \simeq -9/\gamma$$

$$\nabla_{w_1} L(f_{\gamma}, y_{\gamma}) = (f_{\gamma} - y_{\gamma}) u_\gamma = (1/\cdot 1 \times 1 \cdot^{-\delta\cdot} - 1) \times \Lambda/1 \simeq -\Lambda/1$$

$$\Delta_{w_1} L = \frac{1}{\gamma}(-9/\gamma - \Lambda/1) = -\Lambda/V\delta$$

$$w_1 = -\cdot/\cdot \Lambda - (\cdot/\cdot 1)(-\Lambda/V\delta) = \cdot/\cdot\cdot V$$

$$\nabla_{w_\gamma} L(f_{\gamma}, y_{\gamma}) = (f_{\gamma} - y_{\gamma}) \simeq -1$$

$$\nabla_{w_\gamma} L(f_{\gamma}, y_{\gamma}) = (f_{\gamma} - y_{\gamma}) \simeq -1$$

$$\nabla_{w_\gamma} L = \frac{1}{\gamma}(-1 - 1) = \frac{-\gamma}{\gamma} = -1$$

$$\Rightarrow w_\gamma = -\cdot/\cdot\cdot\gamma - (\cdot/\cdot 1)(-1) = \cdot/\cdot\cdot 4$$

$$\omega_0 = 1{,}\mathsf{۴۳۸} \quad , \quad \omega_1 = 0{,}00\mathsf{۷} \quad , \quad \omega_\mathsf{۲} = 0{,}00\mathsf{۴}$$

$$f_1 = \frac{1}{1 + e^{-(1{,}\mathsf{۴۳۸} \times \mathsf{۱۲۱} + 0{,}00\mathsf{۷} \times \mathsf{۱۴}{,}\mathsf{۸} + 0{,}00\mathsf{۴})}} \simeq 1$$

$$f_\mathsf{۲} = \frac{1}{1 + e^{-(1{,}\mathsf{۴۳۸} \times \mathsf{۱۱۴} + \mathsf{۹}00{,}\mathsf{۷} \times \mathsf{۱۵}{,}\mathsf{۲} + 0{,}00\mathsf{۴})}} \simeq 1$$

$$f_\mathsf{۳} = \frac{1}{1 + e^{-(1{,}\mathsf{۴۳۸} \times \mathsf{۲۱}0 + 0{,}00\mathsf{۷} \times \mathsf{۹}{,}\mathsf{۴} + 0{,}00\mathsf{۴})}} \simeq 1$$

$$f_\mathsf{۴} = \frac{1}{1 + e^{-(1{,}\mathsf{۴۳۸} \times \mathsf{۱۹۵} + 0{,}00\mathsf{۷} \times \mathsf{۸}{,}\mathsf{۱} + 0{,}00\mathsf{۴})}} \simeq 1$$

در این تکرار داده های ۴،۳ را نشان می دهیم که خروجی آن ها در گام قبلی به ترتیب به ۰ و ۱ شد در حالیکه برچسب آن ها ۱ می باشد. پس شکل جریمه زیادی را متحمل می شود تا بتواند خروجی فرد را تصحیح کند. این جریمه به قدری زیاد است که شده پارامترهای w را به گمتری سمت به علاوه بر خروجی داده ۴،۳ ، خروجی داده ۱ و ۲ هم ۱ می شوند.

سؤال ۳ ⃝

$$\Delta w_0 l\, (f_1, y_1) = (1-0) \times 121 \simeq 121$$

$$\Delta w_0 l\, (f_4, y_4) = (1-1) \times 195 \simeq 0$$

$$\Delta w_0 l = \frac{1}{2}\left(121 + 0\right) \simeq 4.15$$

$$w_0 = 1.438 - (0.01)(4.15) = 0.833$$

$$\Delta w_1 l\, (f_1, y_1) = (1-0) \times 14.8 \simeq 14.8$$

$$\Delta w_1 l\, (f_4, y_4) = (1-1) \times 8.1 \simeq 0$$

$$\Delta w_1 l = \frac{1}{2}\left(14.8 + 0\right) \simeq 8.4$$

$$w_1 = 0.007 - (0.01)(8.4) = -0.077$$

$$\Delta w_2 l\,(f_1, y_1) = (1-0) = 1$$

$$\Delta w_2 l\,(f_4, y_4) = (1-1) = 0$$

$$\Delta w_2 l = \frac{1}{2}(1+0) = 0.5$$

$$w_2 = 0.009 - (0.01)(0.5) = 0.001$$

$$\omega_0 = 0/\Lambda\Upsilon\Upsilon \; , \; \omega_1 = -0/0VV \; , \; \omega_\Upsilon = 0/001$$

$$f_1 = \cfrac{1}{1+\boxed{e^{-(0/\Lambda\Upsilon\Upsilon \times 1\Upsilon1 - 0/0VV \times 1\Upsilon/\Lambda + 0/001)}}} \simeq 1$$

$$\downarrow$$
$$\Upsilon/1\Upsilon \times 10^{-\Upsilon\Upsilon}$$

$$f_\Upsilon = \cfrac{1}{1+e^{-(0/\Lambda\Upsilon\Upsilon \times 11\Upsilon - 0/0VV \times 10/\Upsilon + 0/001)}} \simeq 1$$

$$f_\Upsilon = \cfrac{1}{1+e^{-(0/\Lambda\Upsilon\Upsilon \times \Upsilon10 - 0/0VV \times 9/\Upsilon + 0/001)}} \simeq 1$$

$$f_\Upsilon = \cfrac{1}{1+e^{-(0/\Lambda\Upsilon\Upsilon \times 19\Delta - 0/0VV \times \Lambda/1 + 0/001)}} \simeq 1$$

در این کسره داده‌ی ۱ در کلاس قطبی خروجی ۱ داشته و برچسب آن ۰

می‌باشد. پس شکل برای این داده حریص می‌شود اما داده‌ی ۴

در کلاس قطبی خروجی ۱ داشته و برچسب هم ۱ می‌باشد پس حریص

نمی‌شود و گرادیان همه‌ی پارامترها برای این داده ۰ می‌شود. از آن طی

به سانتی گرادیان دو داده کافی می‌شود پارامترها خطی تقیم می‌کند و

در نتیجه خروجی جدید داده‌ی ۱ همچنان ۱ مانده و تقیم می‌شود.

$$\Delta_{w_0} L(f_2, y_2) = (f_2 - y_2) u_1 = (1 - 0) \times 114 \simeq 114$$

$$\Delta_{w_0} L(f_3, y_3) = (f_3 - y_3) u_1 = (1 - 1) \times 210 \simeq 0$$

$$\Delta_{w_0} L = \frac{1}{2} (114 + 0) \simeq 57$$

$$w_0 = 0/833 - (0/01)(57) = 0/243$$

$$\Delta_{w_1} L(f_2, y_2) = (f_2 - y_2) u_2 = (1 - 0) \times 15/2 \simeq 15/2$$

$$\Delta_{w_1} L(f_3, y_3) = (f_3 - y_3) u_2 = (1 - 1) \times 9/4 \simeq 0$$

$$\Delta_{w_1} L = \frac{1}{2} (15/2 + 0) = 7/9$$

$$w_1 = -0/077 - (0/01)(7/9) = -0/153$$

$$\Delta_{w_2} L(f_2, y_2) = (f_2 - y_2) \simeq 1$$

$$\Delta_{w_2} L(f_3, y_3) = (f_3 - y_3) \simeq 0$$

$$\Delta_{w_2} L = \frac{1}{2} (1 + 0) = 0/5$$

$$w_2 = 0/001 - (0/01)(0/5) = -0/004$$

$$\omega_0 = 0/۲۴۳ \quad , \quad \omega_1 = -0/۱۵۳ \, , \quad \omega_۲ = -0/۰۰۴$$

$$f_1 = \frac{1}{1 + e^{-(0/۲۴۳ \times ۱۲۱ \, - \, 0/۱۵۳ \times ۱۴/۸ \, - \, 0/۰۰۴)}} \simeq 1$$

$$f_۲ = \frac{1}{1 + e^{-(0/۲۴۳ \times ۱۱۴ \, - \, 0/۱۵۳ \times ۱۵/۲ \, - \, 0/۰۰۴)}} \simeq 1$$

$$f_۳ = \frac{1}{1 + e^{-(0/۲۴۳ \times ۲۱۰ \, - \, 0/۱۵۳ \times ۹/۴ \, - \, 0/۰۰۴)}} \simeq 1$$

$$f_۴ = \frac{1}{1 + e^{-(0/۲۴۳ \times ۱۹۵ \, - \, 0/۱۵۳ \times ۸/۱ \, - \, 0/۰۰۴)}} \simeq 1$$

مانند توضیح قسمت قبلی داده‌ی ۲ هم دقیقاً همین آنساز می‌مانند و

خروجی داده‌ی ۲ تغییر نمی‌شود و درنهایت برحسب بسی بینی شده

برای همین داده‌ها ۱ می‌شود. به نظر من دلیل این آنساز تعداد

کم epoch می‌باشد. اگر تعداد epoch زیاد شود در نهایت

می‌توان به برحسب های صحیح همگرا شد؟ البته می‌توان learning rate

هم زیاد تر کرد تا سریعتر همگرا شد

تا حد معتدل (به قدری زیاد شود که نظر ی min سراسر را

رد کند)