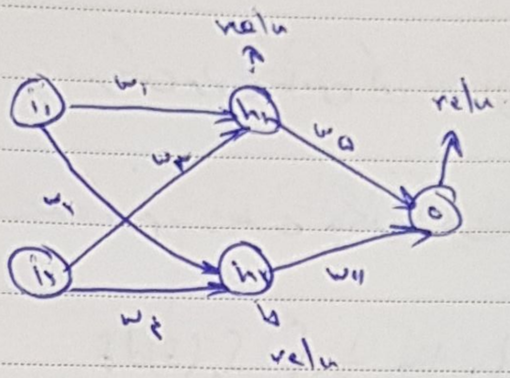
سوال ۲: loss : MSE, optimization : adam, learning rate : ۰,۵



$w_۱ = \frac{1}{[illegible]}$

$w_۲ = -\frac{1}{[illegible]}$

$w_۳ = \frac{1}{[illegible]}$

$w_۴ = -\frac{1}{[illegible]}$

$w_۵ = \frac{1}{[illegible]}$

$w_۶ = -\frac{1}{[illegible]}$

epoch : 1

$X_۱ : i_۱ = ۳ , i_۲ = ۲ \Rightarrow \begin{cases} Z_{h_۱} = \frac{1}{[illegible]} \times ۳ + \frac{1}{[illegible]} \times ۲ = [illegible]\ ۵ \\ Z_{h_۲} = -\frac{1}{۲} \times ۳ + -\frac{1}{۲} \times ۲ = -[illegible] \end{cases}$

$\xrightarrow{relu} 0$

$\Rightarrow Z_o = \frac{1}{۲} \times \overset{۵}{[illegible]}$ [illegible] $= [illegible] > 0$

[illegible]

۲,۵

$X_۲ = i_۱ = ۱۵ , i_۲ = ۱۲ \Rightarrow \begin{cases} Z_{h_۱} = \frac{1}{[illegible]} \times ۱۵ + \frac{1}{[illegible]} \times ۱۲ = \overset{۲۷}{[illegible]} \\ Z_{h_۲} = -\overset{۲۷}{[illegible]} \end{cases}$

$\xrightarrow{relu} 0$

$\Rightarrow Z_o = \overset{۱۳,۵}{[illegible]} > 0$

$\Rightarrow loss = \frac{1}{۲} \sum (Y - \hat{Y})^۲ = \frac{1}{۲} \times ((۸ - \overset{۲,۵}{[illegible]})^۲ + (۲۰ - \overset{۱۳,۵}{[illegible]})^۲) = \frac{[illegible]}{۳۶,۲۵}$ [illegible]

back propagation :

$\frac{d\,loss}{d\hat{y}_i} =$ [illegible] $-۲(y - \hat{y}_i)$

ادامه صفحه بعد ⟶

$[illegible] = \begin{cases} 0 & x < 0 \\ 1 & x > 0 \end{cases} \leftarrow \frac{d\,relu(x)}{d(x)} :$

$$\frac{dloss}{dw_6} = \frac{dloss}{d\hat{y}} \times \underbrace{\frac{d\hat{y}}{dz}}_{(relu)'} \times \underbrace{\frac{dz_0}{dw_6}}_{h_2} = \frac{dloss}{d\hat{y}} \times relu(z_0)' \times h_2$$

$$\frac{dloss}{dw_5} = \frac{dloss}{d\hat{y}} \times \frac{d\hat{y}}{dz} \times \frac{dz_0}{dw_5} = \frac{dloss}{d\hat{y}} \times relu(z_0)' \times h_1$$

$$\frac{dloss}{dw_4} = \frac{dloss}{d\hat{y}} \times relu(z_0)' \times w_6 \times relu(z_{h_2})' \times i_2$$

$$\frac{dloss}{dw_3} = 〃 \times 〃 \times 〃 \times 〃 \times i_1$$

$$\frac{dloss}{dw_2} = 〃 \times 〃 \times w_5 \times relu(z_{h_1})' \times i_2$$

$$\frac{dloss}{dw_1} = 〃 \times 〃 \times 〃 \times 〃 \times i_1$$

حال با استفاده از adam و h۲, پارامتر ها را آپدیت می کنیم:

$beta1 = 0.9$

$beta2 = 0.999$

$lr = 0.01$

$\alpha_{Reg}$ = ضریب

dx تابع گرادیان

$$dx = \begin{bmatrix} \frac{dloss}{dw_1} \\ \frac{dloss}{dw_2} \\ \vdots \\ \frac{dloss}{dw_6} \end{bmatrix} + \alpha \times \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_6 \end{bmatrix}$$

آدام ←

$$moment_{first} = beta1 \times moment_{first} + (1 - beta1) \times dx$$

$$moment_{second} = beta2 \times moment_{second} + (1 - beta2) \times dx^2$$

$$\Rightarrow w = w - lr \times \frac{\frac{moment_{first}}{1-beta_1}}{\sqrt{\frac{moment_{second}}{1-beta_2^t}}}$$

حال با استفاده از محاسبات قبل داریم:

$w_4:$ $\begin{cases} x_1: -۲(۸-۵)\times۱\times \cancel{۲٫۵} = \cancel{۰} \\ x_2: -۲(۲۰-\cancel{۱۲٫۵})\times۱\times\cancel{۰} = \cancel{۰} \end{cases}$ $\frac{dloss}{dw_4} = ۰$

$w_5:$ $\begin{cases} -۲(\cancel{۳})\times۱\times۲٫۵ = \overset{-۲۷٫۵}{\cancel{۵۵}} \\ x_2: -\cancel{۲۹}۱۷۵٫۵ \end{cases}$ میانگین $\Rightarrow \frac{dloss}{dw_5} = \cancel{...} - ۱۰۱$

$w_3:$ $\begin{cases} x_1: -۲\times\overset{۵٫۵}{\cancel{۹}}\times۱\times\cancel{۱}\times۱\times۰\times۲ = ۰ \\ x_2 = ۰ \end{cases}$ $\frac{dloss}{dw_3} = ۰$

$w_6:$ $\begin{cases} x_1: -۲\times۵٫۵\times۱\times۰\times۱\times۰\times۴ = ۰ \\ x_2 = ۰ \end{cases}$ $\frac{dloss}{dw_6} = ۰$

$w_2:$ $\begin{cases} x_1: -۲\times۵٫۵\times۱\times۱\times۱\times۴ = -۴۴ \\ x_2: -۲\times۶٫۵\times۱\times۱\times۱\times۱۲ = -۱۵۶ \end{cases}$ $\frac{dloss}{dw_2} = \frac{-۴۴-۱۵۶}{۲} = -۱۰۰$

$w_1:$ $\begin{cases} x_1: -۲\times۵٫۵\times۱\times۱\times۱\times۳ = -۳۲ \\ x_2: -۲\times۵٫۵\times۱\times۱\times۱\times۱۵ = -۱۸۰ \end{cases}$ $\frac{dloss}{dw_1} = \frac{-۳۲-۱۸۰}{۲} = -۱۰۶٫۵$

Subject:
Year: Month: Day: ( )

$$dx = \frac{۱}{۱۰۰} \times \begin{bmatrix} ۱ \\ -۱ \\ ۱ \\ -۱ \\ ۱ \\ -۱ \end{bmatrix} + \begin{bmatrix} -۱۰۴,۵ \\ -۸۹ \\ ۰ \\ ۰ \\ -۱۰۱,۵ \\ ۰ \end{bmatrix} = \begin{bmatrix} -۱۰۴,۴۹ \\ -۸۹,۰۱ \\ ۰,۰۱ \\ -۰,۰۱ \\ -۱۰۱,۴۹ \\ -۰,۰۱ \end{bmatrix}$$

$$moment_{first} = ۰,۹ \times ۰ + ۰,۱ \times dx = \begin{bmatrix} -۱,۰۴ \\ -۸,۹۰ \\ ۱ \\ -۱ \\ -۱,۰۱ \\ -۱ \end{bmatrix}$$

$$moment_{second} = ۰,۹ \quad ۰,۰۰۱ \times (dx)^۲ = \begin{bmatrix} ۱,۱۳ \\ ۷,۹۲ \\ ۱,۰ \\ ۱,۰ \\ ۱,۰۳ \\ ۱,۰ \end{bmatrix}$$

$$w = w - lr \times \begin{bmatrix} -۱ \\ -۱ \\ ۱ \\ -۱ \\ -۱ \\ -۱ \end{bmatrix} = \begin{bmatrix} ۱,۱ \\ -۰,۹ \\ ۰,۹ \\ -۰,۹ \\ ۱,۱ \\ -۰,۹ \end{bmatrix}$$

epoch ۲:

$$x_۱ = \begin{cases} z_{h_۱} = ۳ \times ۱,۱ + ۲ \times ۰,۹ = ۵,۱ \\ z_{h_۲} = ۳ \times -۰,۹ + ۲ \times -۰,۹ = \text{منفی} \xrightarrow{} relu = ۰ \end{cases} \rightarrow z_o = ۲,۵۴$$

$$x_۲ = \begin{cases} z_{h_۱} = ۱۵ \times ۱,۱ + ۱۲ \times ۰,۹ = ۲۷,۳ \\ z_{h_۲} \overset{relu}{=} ۰ \end{cases} \rightarrow z_o = ۱۳,۴۵$$

$$loss = \frac{۱}{۲}\left((۸ - ۲,۵۴)^۲ + (۲۰ - ۱۳,۴۵)^۲\right) = \boxed{۳۴,۹۵}$$

$$w_۴ : \begin{cases} x_۱ = -۲ \times ۵,۴۴ \times ۰ = ۰ \\ x_۲ = ۰ \end{cases} \Rightarrow \frac{d_{loss}}{d_{w_۴}} = ۰$$

$$w_۵ : \begin{cases} x_۱ = -۲ \times ۵,۴۴ \times ۱ \times ۵,۱ = -۵۵,۴۸ \\ x_۲ = -۲ \times ۶,۳۵ \times ۱ \times ۲۷,۳ = -۳۴۶,۷۱ \end{cases} \quad \boxed{\frac{d_{loss}}{d_{w_۵}} = -۲۰۱,۱}$$

$$\frac{d_{loss}}{d_{w_۳}} = ۰ \quad \frac{d_{loss}}{d_{w_c}} = ۰$$

$$w_۲ : \begin{cases} x_۱ = -۲ \times ۵,۴۴ \times ۱ \times ۱,۱ \times ۱ \times ۲ = -۲۳,۹۳ \\ x_۲ = -۲ \times ۶,۳۵ \times ۱ \times ۱,۱ \times ۱ \times ۱۲ = -۱۶۷,۶۴ \end{cases} \Rightarrow \boxed{\frac{d_{loss}}{d_{w_۲}} = -۹۵,۷۸۵}$$

$$w_۱ \begin{cases} x_۱ : -۳۵,۹ \\ x_۲ : -۲۰۹,۵۵ \end{cases} \Rightarrow \boxed{\frac{d_{loss}}{d_{w_۱}} = -۱۲۲,۷۲}$$

$$dw = epol \times \cancel{loss} + loss = \begin{bmatrix} -۱,۲۲ \\ -۹,۵۷ \\ ۹ \\ -۹ \\ -۲ \\ -۹ \end{bmatrix}$$

$$moment_{first} = ۰,۹ \times \cancel{\ } \, moment_{first} + ۰,۱ \times dw = \begin{bmatrix} -۱۳,۲۲ \\ -۱۷,۵۸ \\ ۰,۹ \\ -۰,۹ \\ -۲۱,۰ \\ -۰,۹ \end{bmatrix}$$

$$m_s = ۰,۹۹۹ \times m_s + ۰,۰۰۱ \times dw^۲ = \begin{bmatrix} ۱۴,۱۸ \\ ۱۷,۰۸ \\ ۰,۹۹ \\ ۰,۹۹ \\ ۴۱,۴۳ \\ ۰,۹۹ \end{bmatrix}$$

$$w = w - lr \times \frac{\frac{m_f}{1-b_1^t}}{\sqrt{\frac{m_s}{1-b_2^t}}} = \begin{bmatrix} ۱,۱۷ \\ -۰,۸۷۹ \\ -۰,۸۷ \\ -۰,۸۷ \\ ۱,۱۷ \\ -۰,۸۷ \end{bmatrix}$$

Subject

Year Month Day ( )

$$x_1 : \begin{cases} z_{h_1} = ۲ \times ۱٫۱۷ + ۳ \times ۰٫۸۷ = \overset{۵٫۲۵}{\cancel{\quad}} \\ z_{h_2} = \text{منفی} \xrightarrow{relu} = ۰ \end{cases} \rightarrow z_o = \overset{۲٫۶۲}{\cancel{\quad}}$$

$$x_2 : \begin{cases} z_{h_1} = ۱۴ \times ۱٫۱۷ + ۱۲ \times ۰٫۸۷ = ۲۷٫۹۸ \\ z_{h_2} \rightarrow ۰ \end{cases} \rightarrow z_o = ۱۳٫۹۹$$

$$\text{loss} = \frac{۱}{۲}\left((۸ - ۲٫۶۲)^۲ + (۲۰ - ۱۳٫۹۹)^۲\right) = \boxed{۳۲٫۵۳}$$

تابع ضرر در ابتدا ۳۶٫۲۵ بعد از ایپوک اول ۳۴٫۹۵ و در آخر برابر ۳۲٫۵۳ بود که نشان از پایین آمدن تابع ضرر و همگرایی دارد و نشان می دهد شبکه دارد خوب تمرین می شود.