CS182 HW3 writing 王鹏豪 2021533138

1. (a) Define the loss funcion is that L. denote that $y=a_5(h_5)$

then use the rule of chain. we have , Define that $\eta\frac{\partial L}{\partial y}\frac{\partial y}{\partial a_5}\frac{\partial a_5}{\partial h_5}$ be λ

as we have

$$\Delta w_{03}=\eta\frac{\partial L}{\partial w_{03}} = \eta\frac{\partial L}{\partial y}\frac{\partial y}{\partial a_5}\frac{\partial a_5}{\partial h_5}\frac{\partial h_5}{\partial a_3}\frac{\partial a_3}{\partial h_3}\frac{\partial h_3}{\partial w_{03}} = \lambda w_{35}\frac{\partial a_3}{\partial h_3} = d_3$$

$$\Delta w_{04}=\eta\frac{\partial L}{\partial w_{04}} = \eta\frac{\partial L}{\partial y}\frac{\partial y}{\partial a_5}\frac{\partial a_5}{\partial h_5}\frac{\partial h_5}{\partial a_4}\frac{\partial a_4}{\partial h_4}\frac{\partial h_4}{\partial w_{04}} = \lambda w_{45}\frac{\partial a_4}{\partial h_4} = d_4$$

then

$$\Delta w_{13}=\eta\frac{\partial L}{\partial w_{13}} = \eta\frac{\partial L}{\partial y}\frac{\partial y}{\partial a_5}\frac{\partial a_5}{\partial h_5}\frac{\partial h_5}{\partial a_3}\frac{\partial a_3}{\partial h_3}\frac{\partial h_3}{\partial w_{13}} = \lambda w_{35}\frac{\partial a_3}{\partial h_3}x_1 = \lambda w_{35}\frac{\partial a_3}{\partial h_3} = d_3$$

$$\Delta w_{23}=\eta\frac{\partial L}{\partial w_{23}} = \eta\frac{\partial L}{\partial y}\frac{\partial y}{\partial a_5}\frac{\partial a_5}{\partial h_5}\frac{\partial h_5}{\partial a_3}\frac{\partial a_3}{\partial h_3}\frac{\partial h_3}{\partial w_{23}} = \lambda w_{35}\frac{\partial a_3}{\partial h_3}x_2 = -\lambda w_{35}\frac{\partial a_3}{\partial h_3} = -d_3$$

$$\Delta w_{14}=\eta\frac{\partial L}{\partial w_{14}} = \eta\frac{\partial L}{\partial y}\frac{\partial y}{\partial a_5}\frac{\partial a_5}{\partial h_5}\frac{\partial h_5}{\partial a_4}\frac{\partial a_4}{\partial h_4}\frac{\partial h_4}{\partial w_{14}} = \lambda w_{45}\frac{\partial a_4}{\partial h_4}x_1 = \lambda w_{45}\frac{\partial a_4}{\partial h_4} = d_4$$

$$\Delta w_{24}=\eta\frac{\partial L}{\partial w_{24}} = \eta\frac{\partial L}{\partial y}\frac{\partial y}{\partial a_5}\frac{\partial a_5}{\partial h_5}\frac{\partial h_5}{\partial a_4}\frac{\partial a_4}{\partial h_4}\frac{\partial h_4}{\partial w_{24}} = \lambda w_{45}\frac{\partial a_4}{\partial h_4}x_2 = -\lambda w_{45}\frac{\partial a_4}{\partial h_4} = -d_4$$

So, we have that

$$\Delta w_{13}=d_3,\quad \Delta w_{23}=-d_3,\quad \Delta w_{14}=d_4 \quad \Delta w_{24}=-d_4$$

(b) Firstly, as we have

$$h_3=w_{03}+w_{13}x_1+w_{23}x_2,\quad h_4=w_{04}+w_{14}x_1+w_{24}x_2$$

$$h_5=w_{05}+w_{35}a_3(h_3)+w_{45}a_4(h_4)$$

then we get $h_3=-1+1\times1+(-1)\times(-1)=1$, $h_4=2+(-1)\times1+1\times(-1)=0$

$g(h_3)=\frac{1}{1+e^{-1}}$, $g(h_4)=\frac{1}{1+e^0}=\frac{1}{2}$, $h_5=1\times\frac{1}{1+e^{-1}}+1\times\frac{1}{2}-2=\frac{1}{1+e^{-1}}-\frac{3}{2}$

$y=g(h_5)=\frac{1}{1+e^{-h_5}}=0.317$. loss: $(\hat{y}-y)^2=(0.317-1)^2=0.467$

so we get that after forward propagation, the output is 0.317. error is 0.467

then we make back propagation

$$\frac{\partial L}{\partial w_{03}}=\frac{\partial L}{\partial y}\frac{\partial y}{\partial a_5}\frac{\partial a_5}{\partial h_5}\frac{\partial h_5}{\partial a_3}\frac{\partial a_3}{\partial h_3}\frac{\partial h_3}{\partial w_{03}}=2(y-1)\cdot1\cdot g(h_5)(1-g(h_5))\cdot w_{35}\cdot g(h_3)[1-g(h_3)]\cdot1$$

as $y=g(h_5)=0.317$. we have $\frac{\partial L}{\partial w_{03}}=-0.058$

$$\frac{\partial L}{\partial w_{04}}=\frac{\partial L}{\partial y}\frac{\partial y}{\partial a_5}\frac{\partial a_5}{\partial h_5}\frac{\partial h_5}{\partial a_4}\frac{\partial a_4}{\partial h_4}\frac{\partial h_4}{\partial w_{04}}=2(y-1)\cdot1\cdot g(h_5)(1-g(h_5))\cdot w_{45}\, g(h_4)[1-g(h_4)]\cdot1$$

as $y=g(h_5)=0.317$. we have $\frac{\partial L}{\partial w_{04}}=-0.074$

$\frac{\partial L}{\partial w_{05}}=\frac{\partial L}{\partial y}\frac{\partial y}{\partial a_5}\frac{\partial a_5}{\partial h_5}\frac{\partial h_5}{\partial w_{05}}=2(y-1)\cdot1\cdot g(h_5)(1-g(h_5))=-0.296$, using (a), as $x_1=1$, $x_2=-1$

we get $\frac{\partial L}{\partial w_{13}}=d_3=-0.058$ $\frac{\partial L}{\partial w_{23}}=-d_3=0.058$, $\frac{\partial L}{\partial w_{14}}=d_4=-0.074$, $\frac{\partial L}{\partial w_{24}}=-d_4=0.074$

then, as for $\frac{\partial L}{\partial w_{35}}$, $\frac{\partial L}{\partial w_{45}}$, we have

$$\frac{\partial L}{\partial w_{35}} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial a_5}\frac{\partial a_5}{\partial h_5}\frac{\partial h_5}{\partial w_{35}} = 2(y-1)\cdot 1\cdot g(h_5)[1-g(h_5)]\, g(h_3) = -0.216$$

$$\frac{\partial L}{\partial w_{45}} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial a_5}\frac{\partial a_5}{\partial h_5}\frac{\partial h_5}{\partial w_{45}} = 2(y-1)\cdot 1\cdot g(h_5)[1-g(h_5)]\, g(h_4) = -0.148$$

so we get that, the new weight and bias are:

$w'_{03} = w_{03} - \eta\frac{\partial L}{\partial w_{03}} = -0.971$, $w'_{04} = w_{04} - \eta\frac{\partial L}{\partial w_{04}} = 2.037$, $w'_{05} = w_{05} - \eta\frac{\partial L}{\partial w_{05}} = -1.852$

$w'_{13} = w_{13} - \eta\frac{\partial L}{\partial w_{13}} = 1.029$, $w'_{14} = w_{14} - \eta\frac{\partial L}{\partial w_{14}} = -0.963$, $\cancel{w'_{15} = w_{15} - \eta\frac{\partial L}{\partial w_{15}} =}$

$w'_{23} = w_{23} - \eta\frac{\partial L}{\partial w_{23}} = -1.029$, $w'_{24} = w_{24} - \eta\frac{\partial L}{\partial w_{24}} = 0.963$.

$w'_{35} = \cancel{110}\ w_{35} - \eta\frac{\partial L}{\partial w_{35}} = 1.108$, $w'_{45} = w_{45} - \eta\frac{\partial L}{\partial w_{45}} = 1.074$

Then again apply forward propagation.

$y' = g(h'_5) = 0.388$ error: $(y'-1)^2 = 0.375$

2.(a)

i 2

ii 0.881

iii 8

iv 0.999665

v 3.643

vi 0.262

vii. 3

viii. 1.341

(b)

i. ~~1.651~~ 1.6505

ii. ~~0.892~~ 0.8918

iii. ~~2.000~~ 2.0000

iv. ~~0.999~~ 0.9986

v. 2

(c)

i.

$$\ell'(\hat{y},y) = \ell(\hat{y},y) + \frac{\lambda}{2}(\|\alpha\|_2^2 + \|\beta\|_2^2) = -\sum_{i=1}^{3} y_i \log(\hat{y}_i) + \frac{\lambda}{2}\|\alpha\|_2^2 + \frac{\lambda}{2}\|\beta\|_2^2$$

$$\cancel{\ell'(\hat{y},y) = \ell(\hat{y},y) + \lambda(\|\alpha\|_2^2 + \|\beta\|_2^2) = \ell(\hat{y},y) + \lambda\left[\sum_{j,i}(\alpha_{j,i})^2 + \sum_{j,i}(\beta_{j,i})^2\right]}$$

$$\cancel{= \ell(\hat{y},y) + \lambda\left(\sum_{j,i}\alpha_{j,i}^2 + \sum_{j,i}\beta_{j,i}^2\right)}$$

$$\cancel{\ell'(\hat{y},y) = \ell(\hat{y},y) + \lambda(\|\alpha\|_2^2)}$$

$$\cancel{= \ell(\hat{y},y) + \lambda\sum_{j=1}(\alpha_{j,i})^2}$$

ii. ~~2.441~~ 1.8762

iii. ~~1.408~~ 1.4076

iv. ~~17.698~~ 40.5526

v. ~~grd~~ gradient update equation

$$\alpha_{j,i} = \alpha_{j,i} - \eta\left(\frac{\partial \ell(\hat{y},y)}{\partial \alpha_{j,i}} + \lambda \alpha_{j,i}\right)$$

vi. the second and the fourth are true

3\. (a) Here we use cross entropy as the loss function as we use logistic regression

$$l = \frac{1}{L}\sum_{l=1}^{L}[-y_l\log(\hat{y}_l) - (1-y_l)\log(1-\hat{y}_l)],$$

we then make derivation

$$\frac{\partial l}{\partial \hat{y}_l} = \frac{1}{L}\left(-y_l\frac{1}{\hat{y}_l} + (1-y_l)\frac{1}{1-\hat{y}_l}\right)$$

, as for ~~what~~ $l' \neq l$. the item is zero

as we have that $\hat{y}_l = \mathrm{sigmoid}(w^T x_l)$, then for w_i in w, we have

so $\frac{\partial l}{\partial w_i} = \sum_{l=1}^{L}\frac{\partial l}{\partial \hat{y}_l}\frac{\partial \hat{y}_l}{\partial w^T x_l}\frac{\partial w^T x_l}{\partial w_i}$. as we have for sigmoid $f'(x) = f(x)(1-f(x))$

$$= \sum_{l=1}^{L}\frac{1}{L}\left(-y_l\frac{1}{\hat{y}_l} + (1-y_l)\frac{1}{1-\hat{y}_l}\right)\hat{y}_l(1-\hat{y}_l)x_{l,i}$$

, as x is a 2D var

$$= \sum_{l=1}^{L}\frac{1}{L}(-y_l(1-\hat{y}_l) + (1-y_l)\hat{y}_l)x_{l,i}$$

$$= \frac{1}{L}\sum_{l=1}^{L}(\hat{y}_l - y_l)x_{l,i}$$

then we have

$$w_i' = w_i - \alpha\frac{\partial l}{\partial w_i} = w_i - \frac{\alpha}{L}\sum_{l=1}^{L}(\hat{y}_l - y_l)x_{l,i}$$

(b) then as for $P(y_l=1|\hat{y}_l) = \mathrm{sigmoid}(w^T x_l)$ and $P(y_l=-1|\hat{y}_l) = \mathrm{sigmoid}(-w^T x_l)$

we make function from the y_l in (b) to y_l in (a)

as $y_l \in \{-1,1\}$, to $y_{la} \in \{0,1\}$

that is

$y_l = 2y_{la} - 1 \quad y_{la} = \frac{1+y_l}{2}$

so $w_i' = w_i - \alpha\frac{\partial l}{\partial w_i} = w_i - \frac{\alpha}{L}\sum_{l=1}^{L}\left(\hat{y}_l - \frac{1+y_l}{2}\right)x_{l,i}$

$$\Rightarrow w_i' = w_i - \frac{\alpha}{L}\sum_{l=1}^{L}\left(\hat{y}_l - \frac{y_l}{2} - \frac{1}{2}\right)x_{l,i}$$