Algebros uždaviniai

1. Duotos vienodos dimensijos matricos $\boldsymbol{A}, \boldsymbol{B}$ ir $\boldsymbol{I}$. Apskaičiuokite matricą $\boldsymbol{C}$ (0.05 t.):

$$\boldsymbol{C} = \boldsymbol{A} \odot \boldsymbol{I} + \boldsymbol{A}\boldsymbol{B}, \quad \boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} -1 & 1 \\ 4 & 4 \end{pmatrix} \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$C = \begin{pmatrix} 1\cdot 1 & 2\cdot 0 \\ 2\cdot 0 & 3\cdot 1 \end{pmatrix} + \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} -1 & 1 \\ 4 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} + \begin{pmatrix} 7 & 9 \\ 10 & 14 \end{pmatrix} = \boxed{\begin{pmatrix} 8 & 9 \\ 10 & 17 \end{pmatrix}}$$

2. Raskite $\boldsymbol{Y} = \boldsymbol{K} * \boldsymbol{A}$, kai operacija atliekama su padinimu nuliais (angl. padding) $p = 2$, bei žingsniu (angl. stride) $s = 2$ (0.1 t.):

$$\boldsymbol{A} = \begin{pmatrix} 3 & 3 & 2 & 1 & 0 \end{pmatrix}, \quad \boldsymbol{K} = \begin{pmatrix} -1 & 2 & -1 \end{pmatrix}$$

$\downarrow p=2$

$$A_{p=2} = \begin{pmatrix} 0 & 0 & 3 & 3 & 2 & 1 & 0 & 0 & 0 \end{pmatrix}$$

$$\boxed{Y = \begin{pmatrix} -3 & 1 & 0 & 0 \end{pmatrix}}$$

3. Raskite $\boldsymbol{Y} = \boldsymbol{K} * \boldsymbol{A}$, kai operacija atliekama su padinimu nuliais (angl. padding) $p = (2, 0)$, bei žingsniu (angl. stride) $s = (2, 1)$ (0.2 t.):

$$\boldsymbol{A} = \begin{pmatrix} 3 & 3 & 2 & 1 & 0 \\ 0 & 0 & 1 & 3 & 1 \\ 3 & 1 & 2 & 2 & 3 \\ 2 & 0 & 0 & 2 & 2 \\ 2 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{K} = \begin{pmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{pmatrix}$$

$\downarrow p=(2,0)$

$$A_{p=(2,0)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 3 & 3 & 2 & 1 & 0 \\ 0 & 0 & 1 & 3 & 1 \\ 3 & 1 & 2 & 2 & 3 \\ 2 & 0 & 0 & 2 & 2 \\ 2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \boxed{Y = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 0 & 3 \\ -7 & -1 & 0 \\ -2 & 0 & -1 \end{pmatrix}}$$

4. Raskite $\mathbf{Y} = \mathbf{K}_{3\times 2\times 3\times 3} * \mathbf{A}_{2\times 3\times 3}$, kai operacija atliekama su padinimu nuliais (angl. padding) $p = (0, 0, 0)$, bei žingsniu (angl. stride) $s = (1, 1, 1)$ (0.4 t.):

$$\mathbf{A} = (\boldsymbol{A}_1, \boldsymbol{A}_2) = \left(\begin{pmatrix} 3 & 3 & 2 & 1 & 0 \\ 0 & 0 & 1 & 3 & 1 \\ 3 & 1 & 2 & 2 & 3 \\ 2 & 0 & 0 & 2 & 2 \\ 2 & 0 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \right) \qquad Y = (Y_1, Y_2, Y_3)$$

$$\mathbf{K} = (\mathbf{K}_1, \mathbf{K}_2, \mathbf{K}_3)$$

$$\mathbf{K}_1 = (\boldsymbol{K}_{1,1}, \boldsymbol{K}_{1,2}) = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right) \qquad Y_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{K}_2 = (\boldsymbol{K}_{2,1}, \boldsymbol{K}_{2,2}) = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right) \qquad Y_2 = \begin{pmatrix} 0 & 1 & 3 \\ 1 & 2 & 2 \\ 0 & 0 & 2 \end{pmatrix}$$

$$\mathbf{K}_3 = (\boldsymbol{K}_{3,1}, \boldsymbol{K}_{3,2}) = \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right) \qquad Y_3 = \begin{pmatrix} 4 & 4 & 5 \\ 1 & 3 & 3 \\ 3 & 1 & 4 \end{pmatrix}$$

$$\boxed{Y = \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 3 \\ 1 & 2 & 2 \\ 0 & 0 & 2 \end{pmatrix}, \begin{pmatrix} 4 & 4 & 5 \\ 1 & 3 & 3 \\ 3 & 1 & 4 \end{pmatrix} \right)}$$

Tikimybių teorijos/statistikos uždaviniai

1. Tarkime turime žymes $\boldsymbol{Y}$ ir modelio prognozes $\hat{\boldsymbol{Y}}$. Raskite klafifikavimo lenteles kiekvienos klasės atžvilgniu. Raskite tikslumo, precizijos ir atkūrimo statistikas antros klasės atžvilgiu (0.2 t.).

Assume multi-class, single label; assume argmax takes lower index if tie

$$\boldsymbol{Y} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \hat{\boldsymbol{Y}} = \begin{pmatrix} 0.2 & 0.8 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \\ 0.01 & 0.99 & 0 \\ 0.99 & 0.01 & 0 \\ 0.4 & 0.3 & 0.3 \\ 0.4 & 0.3 & 0.3 \\ 0.4 & 0.4 & 0.2 \\ 0 & 0.1 & 0.9 \\ 0.1 & 0 & 0.9 \end{pmatrix} \qquad \hat{Y}' = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

	$\hat{Y}=1$	$\hat{Y}=2$	$\hat{Y}=3$
$Y=1$	3	0	0
$Y=2$	1	3	1
$Y=3$	0	0	2

Class 2:

$$Acc = \frac{3+5}{3+5+2} = \frac{8}{10} = 80\%$$

$$Pr = \frac{3}{3+0} = 1 = 100\%$$

$$Re = \frac{3}{3+2} = \frac{3}{5} = 60\%$$

	$\hat{Y}=2$	$\hat{Y}\neq 2$	
$Y=2$	3	2	$\leftarrow$ Re
$Y\neq 2$	0	5	
	$\uparrow$ Pr		

2. Paskaičiuokite makro-F1 statistika (angl. macro-F1) tarp $\boldsymbol{Y}$ ir $\hat{\boldsymbol{Y}}$ (0.2 t.)

$$\text{Macro F1} = \frac{1}{C}\sum_i^C F1_i \qquad F1_i = 2\,\frac{Pr_i \cdot Re_i}{Pr_i + Re_i}$$

$$Pr_1 = \frac{3}{4} \quad Re_1 = \frac{3}{3} \quad F1_1 = 2\,\frac{\frac{3}{4}\cdot 1}{\frac{3}{4}+\frac{4}{4}} = \frac{6}{7}$$

$$Pr_2 = \frac{3}{3} \quad Re_2 = \frac{3}{5} \quad F1_2 = 2\,\frac{1\cdot\frac{3}{5}}{\frac{5}{5}+\frac{3}{5}} = \frac{6}{8}$$

$$Pr_3 = \frac{2}{3} \quad Re_3 = \frac{2}{2} \quad F1_3 = 2\,\frac{\frac{2}{3}\cdot 1}{\frac{2}{3}+\frac{3}{3}} = \frac{4}{5}$$

$$\text{MacroF1} = \frac{1}{3}\left(\frac{6}{7} + \frac{6}{8} + \frac{4}{5}\right) = \frac{2}{7} + \frac{2}{8} + \frac{4}{15} \approx \boxed{0.8}$$

3. Paskaičiuokite kryžminės entropijos nuostolius tarp $\boldsymbol{Y}$ ir $\hat{\boldsymbol{Y}}$ (0.15 t.)

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_i^N \sum_j^C y_{ij} \log \hat{y}_{ij}$$

$$= -\frac{1}{10}\big(\log 0.8 + \log 0.5 + \log 0.25 + \log 0.99 + \log 0.01 + \log 0.4 + \log 0.4 + \log 0.4 + \log 0.9 + \log 0.9\big) \approx \boxed{0.988}$$

$$\boldsymbol{Y} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \hat{\boldsymbol{Y}} = \begin{pmatrix} 0.2 & 0.8 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \\ 0.01 & 0.99 & 0 \\ 0.99 & 0.01 & 0 \\ 0.4 & 0.3 & 0.3 \\ 0.4 & 0.3 & 0.3 \\ 0.4 & 0.4 & 0.2 \\ 0 & 0.1 & 0.9 \\ 0.1 & 0 & 0.9 \end{pmatrix}$$

Tarkime parametrus vertiname Nestarovo metodu. Apskaičiuokite $\boldsymbol{\theta}^{(2)}$, jei duotos pradinės reikšmės (0.3 t.):

τ	$\nabla\mathcal{L}_{\boldsymbol{\theta}}^{(\tau)}$	$\boldsymbol{\theta}^{(\tau)}$	$\boldsymbol{v}^{(\tau)}$
0	$\begin{pmatrix}10\\5\end{pmatrix}$	$\begin{pmatrix}1\\1\end{pmatrix}$	$\begin{pmatrix}1\\1\end{pmatrix}$
1	$\begin{pmatrix}-4\\2\end{pmatrix}$		
2	$\begin{pmatrix}2\\-1\end{pmatrix}$		

čia $\eta = 0.1, \rho = 0.9.$

$$\boldsymbol{\theta}^{(\tau)} = \boldsymbol{\theta}^{(\tau-1)} - \eta \cdot \Delta\boldsymbol{v}^{(\tau)},$$

$$\begin{cases} \hat{\boldsymbol{\theta}}^{(\tau)} & = \boldsymbol{\theta}^{(\tau-1)} - \eta \cdot \rho \cdot \Delta\boldsymbol{v}^{(\tau-1)}, \quad ? \\ \Delta\boldsymbol{v}^{(\tau)} & = \rho \cdot \Delta\boldsymbol{v}^{(\tau-1)} + (1-\rho) \cdot \underbrace{\nabla_{\boldsymbol{\theta}}\mathcal{L}(\hat{\boldsymbol{\theta}}^{(\tau)})}_{\approx \nabla_\theta \mathcal{L}^{(\tau)}}. \end{cases}$$

$$\theta^{(1)} = \theta^{(0)} - \eta \cdot \Delta v^{(1)}$$

$$\Delta v^{(1)} = \rho \cdot \Delta v^{(0)} + (1-\rho) \cdot \nabla_\theta \mathcal{L}^{(1)}$$

$$= 0.9 \cdot \begin{pmatrix}1\\1\end{pmatrix} + (1-0.9) \cdot \begin{pmatrix}-4\\2\end{pmatrix} = \begin{pmatrix}0.9\\0.9\end{pmatrix} + \begin{pmatrix}-0.4\\0.2\end{pmatrix} = \begin{pmatrix}0.5\\1.1\end{pmatrix}$$

$$\theta^{(1)} = \begin{pmatrix}1\\1\end{pmatrix} - 0.1\begin{pmatrix}0.5\\1.1\end{pmatrix} = \begin{pmatrix}0.95\\0.89\end{pmatrix}$$

τ	$\nabla_\theta \mathcal{L}^{(\tau)}$	$\theta^{(\tau)}$	$\Delta v^{(\tau)}$
0	$\begin{pmatrix}10\\5\end{pmatrix}$	$\begin{pmatrix}1\\1\end{pmatrix}$	$\begin{pmatrix}1\\1\end{pmatrix}$
1	$\begin{pmatrix}-4\\2\end{pmatrix}$	$\begin{pmatrix}0.95\\0.89\end{pmatrix}$	$\begin{pmatrix}0.5\\1.1\end{pmatrix}$
2	$\begin{pmatrix}2\\1\end{pmatrix}$		

$$\Delta v^{(2)} = \rho \cdot \Delta v^{(1)} + (1-\rho)\nabla_\theta \mathcal{L}^{(2)}$$

$$= 0.9\begin{pmatrix}0.5\\1.1\end{pmatrix} + (1-0.9)\begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}0.45\\0.99\end{pmatrix} + \begin{pmatrix}0.2\\0.1\end{pmatrix} = \begin{pmatrix}0.65\\1.09\end{pmatrix}$$

$$\theta^{(2)} = \begin{pmatrix}0.95\\0.89\end{pmatrix} - 0.1\begin{pmatrix}0.65\\1.09\end{pmatrix} = \begin{pmatrix}0.885\\0.781\end{pmatrix}$$

τ	$\nabla_\theta \mathcal{L}^{(\tau)}$	$\theta^{(\tau)}$	$\Delta v^{(\tau)}$
0	$\begin{pmatrix}10\\5\end{pmatrix}$	$\begin{pmatrix}1\\1\end{pmatrix}$	$\begin{pmatrix}1\\1\end{pmatrix}$
1	$\begin{pmatrix}-4\\2\end{pmatrix}$	$\begin{pmatrix}0.95\\0.89\end{pmatrix}$	$\begin{pmatrix}0.5\\1.1\end{pmatrix}$
2	$\begin{pmatrix}2\\1\end{pmatrix}$	$\boxed{\begin{pmatrix}0.885\\0.781\end{pmatrix}}$	$\begin{pmatrix}0.65\\1.09\end{pmatrix}$

2 Kompleksinė užduotis (1 b.)

1. Nagrinėjame paveikslėlių klasifikavimą į tris klases.

   Mūsų turimi duomenys susideda iš ~~dviejų~~ trijų pavyzdžių t.y. matricų $X^{(1)}_{3\times3}$, $X^{(2)}_{3\times3}$ ir $X^{(3)}_{3\times3}$, kurių pirmos dvi atitinkamai priklauso pirmai klasei, trečias pavyzdys antrai klasei ir trečios klasės pavyzdžių neturime.

$$X^{(1)}_{3\times3}=\begin{pmatrix}a&b&0\\a&b&0\\a&b&0\end{pmatrix},\quad X^{(2)}_{3\times3}=\begin{pmatrix}0&c&a\\0&c&a\\0&c&a\end{pmatrix},\quad X^{(3)}_{3\times3}=\begin{pmatrix}a&b&c\\c&b&a\\a&-a&a\end{pmatrix}$$

$y^{(1)}=$ Kl. 1 Kl.3 nėra
$y^{(2)}=$ Kl. 1
$y^{(3)}=$ Kl. 2

Pvz 567

   Čia a,b,c – skaitmenys iš Jūsų paskutinių triejų LSP pažymėjimo nr.

   Nagrinėjamas modelis yra:

$$\begin{aligned}z_1&=pool^{avg}_{3\times3}(relu(K*X)+E)\\z_2&=pool^{avg}_{3\times3}(relu(K*X)+E)\\z_3&=pool^{avg}_{3\times3}(relu(K*X)+E)\end{aligned}$$

$relu(x)=max(0,x)$

$z_1=z_2=z_3$?

$$\begin{aligned}t_1&=w_1z_1+b_1\\t_2&=w_2z_2+b_2\\t_3&=w_3z_3+b_3\\\hat{y}_i&=softmax(t_i),\quad i=1,2,3.\end{aligned}$$

$E=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}$

$softmax(\vec{t})=\frac{e^{t_i}}{\sum_i e^{t_i}}$ $\left(\vec{t}=(t_1,t_2,\dots)\right)$

   čia E - vienetinė diagonalinė matrica.

   Modelio parametrų fiksuotos reikšmės $w_1=1, b_1=-1, w_2=10, b_2=-20, w_3=-10, b_3=10$. Bei $K_{3\times3}$ yra nežinomi parametrai su pradinėmis reikšmėmis:

$$K=\begin{pmatrix}1&0&-1\\-1&2&-1\\1&-2&1\end{pmatrix}$$

   Kompleksinė užduotis prašo paskaičiuoti visus modelio skaičiavimo etapus: tiesioginį apskaičiavimą, gradientų apskaičiavimą, atgalinės sklaidos algoritmą, bei parametrų optimizavimą).

2. 1. (0.1 t.) Užrašykite kokios yra tikros žymės matriciniu pavidalu: yra 3 klasės, todėl matricos plotis 3

   $y^{(1)}=?$
   $y^{(2)}=?$
   $y^{(3)}=?$

$y^{(1)}=(1\ 0\ 0)$
$y^{(2)}=(1\ 0\ 0)$
$y^{(3)}=(0\ 1\ 0)$

$$Y=\begin{pmatrix}1&0&0\\1&0&0\\0&1&0\end{pmatrix}$$

   2. (0.3 t.) Apskaičiuokite $\hat{y}_i$ reikšmes, kai konvoliucija $K*X$, atliekama su papildymu nuliais $p=(1,1)$

pvz. $(a,b,c)=(5,6,7)$

(1)

$$X^{(1)}_{p=(1,1)}=\begin{pmatrix}0&0&0&0&0\\0&5&6&0&0\\0&5&6&0&0\\0&5&6&0&0\\0&0&0&0&0\end{pmatrix}$$

$$K*X^{(1)}_{p=(1,1)}=\begin{pmatrix}0&0&0\\-6&5&6\\-2&12&0\end{pmatrix}$$

$$relu(K*X^{(1)}_{p=(1,1)})=\begin{pmatrix}0&0&0\\0&5&6\\0&12&0\end{pmatrix}$$

$$relu(K*X^{(1)}_{p=(1,1)})+E=\begin{pmatrix}1&0&0\\0&6&6\\0&12&1\end{pmatrix}$$

$$z=pool^{avg}_{3\times3}(relu(K*X^{(1)}_{p=(1,1)})+E)=2\tfrac{8}{9}$$

$$t_1=1\cdot2\tfrac{8}{9}+(-1)=1\tfrac{8}{9}$$
$$t_2=10\cdot2\tfrac{8}{9}+(-20)=8\tfrac{8}{9}$$
$$t_3=(-10)\cdot2\tfrac{8}{9}+10=-18\tfrac{8}{9}$$
$$\hat{y}^{(1)}=softmax\left(\left(1\tfrac{8}{9}\quad 8\tfrac{8}{9}\quad -18\tfrac{8}{9}\right)\right)=$$
$$\approx(0.0009..\quad 0.9990..\quad 0)$$

(2)

$$X^{(2)}_{p=(1,1)}=\begin{pmatrix}0&0&0&0&0\\0&0&7&5&0\\0&0&7&5&0\\0&0&7&5&0\\0&0&0&0&0\end{pmatrix}$$

$$K*X^{(2)}_{p=(1,1)}=\begin{pmatrix}0&0&0\\-7&-5&7\\0&4&10\end{pmatrix}$$

$$relu(K*X^{(2)}_{p=(1,1)})=\begin{pmatrix}0&0&0\\0&0&7\\0&4&10\end{pmatrix}$$

$$relu(K*X^{(2)}_{p=(1,1)})+E=\begin{pmatrix}1&0&0\\0&1&7\\0&4&11\end{pmatrix}$$

$$z=pool^{avg}_{3\times3}(relu(K*X^{(2)}_{p=(1,1)})+E)=2\tfrac{6}{9}$$

$$t_1=1\cdot2\tfrac{6}{9}+(-1)=1\tfrac{6}{9}$$
$$t_2=10\cdot2\tfrac{6}{9}+(-20)=6\tfrac{6}{9}$$
$$t_3=(-10)\cdot2\tfrac{6}{9}+10=-16\tfrac{6}{9}$$
$$\hat{y}^{(2)}=softmax\left(\left(1\tfrac{6}{9}\quad 6\tfrac{6}{9}\quad -16\tfrac{6}{9}\right)\right)$$
$$\approx(0.006..\quad 0.993..\quad 0)$$

(3)

$$X^{(3)}_{p=(1,1)}=\begin{pmatrix}0&0&0&0&0\\0&5&6&7&0\\0&7&6&5&0\\0&5&-5&5&0\\0&0&0&0&0\end{pmatrix}$$

$$K*X^{(3)}_{p=(1,1)}=\begin{pmatrix}-4&0&4\\-13&-2&-10\\-1&-18&21\end{pmatrix}$$

$$relu(K*X^{(3)}_{p=(1,1)})=\begin{pmatrix}0&0&4\\0&0&0\\0&0&21\end{pmatrix}$$

$$relu(K*X^{(3)}_{p=(1,1)})+E=\begin{pmatrix}1&0&4\\0&1&0\\0&0&22\end{pmatrix}$$

$$z=pool^{avg}_{3\times3}(relu(K*X^{(3)}_{p=(1,1)})+E)=3\tfrac{1}{9}$$

$$t_1=1\cdot3\tfrac{1}{9}+(-1)=2\tfrac{1}{9}$$
$$t_2=10\cdot3\tfrac{1}{9}+(-20)=11\tfrac{1}{9}$$
$$t_3=(-10)\cdot3\tfrac{1}{9}+10=-21\tfrac{1}{9}$$
$$\hat{y}^{(3)}=softmax\left(\left(2\tfrac{1}{9}\quad 11\tfrac{1}{9}\quad -21\tfrac{1}{9}\right)\right)$$
$$\approx(0.0001..\quad 0.9998..\quad 0)$$

$$\hat{Y}\approx\begin{pmatrix}0&0.99&0\\0&0.99&0\\0&0.99&0\end{pmatrix}$$

   3. (0.1 t.) Apskaičiuokite kryžminės entropijos nuostolius iš modelio prognozių $\hat{y}$ ir tikrų žymių y.

$$\mathcal{L}_{CE}=-\frac{1}{N}\sum_i^N\sum_j^C y_j^{(i)}\log\hat{y}_j^{(i)}$$

$$Y=\begin{pmatrix}1&0&0\\1&0&0\\0&1&0\end{pmatrix}\quad\hat{Y}=\begin{pmatrix}0.0009&0.999&0\\0.006&0.993&0\\0.0001&0.9998&0\end{pmatrix}$$

$$\approx-\frac{1}{3}\Big(1\cdot\log(0.0009)+0\cdot\log(0.999)+0\cdot\log(0)+$$
$$+1\cdot\log(0.006)+0\cdot\log(0.999)+0\cdot\log(0)+$$
$$+0\cdot\log(0.0001)+1\cdot\log(0.9998)+0\cdot\log(0)\Big)\approx-\frac{1}{3}\big((-7)+(-5)+(0)\big)\approx4$$

   4. (0.3 t.) Apskaičiuokite nuostolių funkcijos išvestines nežinomų parametrų atžvilgiu:

$$\frac{\partial\mathcal{L}}{\partial k_{i,j}},\quad\forall i,j=1,2,3$$

$$\frac{\partial\mathcal{L}}{\partial k}=\frac{\partial\mathcal{L}}{\partial\dots}\cdots\frac{\partial\dots}{\partial k}$$ ← product rule

$$z=pool^{avg}_{3\times3}\underbrace{(relu(\underbrace{K*X}_{c})+E)}_{r}$$

$$\frac{\partial\mathcal{L}}{\partial k}=\frac{\partial\mathcal{L}}{\partial\hat{y}}\frac{\partial\hat{y}}{\partial t}\frac{\partial t}{\partial z}\frac{\partial z}{\partial r}\frac{\partial r}{\partial c}\frac{\partial c}{\partial k}$$

$$\frac{\partial\mathcal{L}}{\partial\hat{y}_j}=-\frac{1}{N}\sum_i^N\frac{\partial}{\partial\hat{y}_j}\left[y_j^{(i)}\log\hat{y}_j^{(i)}\right]$$

$$\frac{\partial\mathcal{L}}{\partial\hat{y}}=\begin{bmatrix}\frac{\partial\mathcal{L}}{\partial\hat{y}_1}&\frac{\partial\mathcal{L}}{\partial\hat{y}_2}&\frac{\partial\mathcal{L}}{\partial\hat{y}_3}\end{bmatrix}=\begin{bmatrix}-\frac{1}{N}\sum_i^N\frac{y_1^{(i)}}{\hat{y}_1^{(i)}}&-\frac{1}{N}\sum_i^N\frac{y_2^{(i)}}{\hat{y}_2^{(i)}}&-\frac{1}{N}\sum_i^N\frac{y_3^{(i)}}{\hat{y}_3^{(i)}}\end{bmatrix}$$

$$\frac{\partial\hat{y}_i}{\partial t_j}=\cancel{\frac{\partial}{\partial t_j}\frac{e^{t_i}}{\sum_{i=1}^{3}e^{t_i}}}$$

$$\frac{\partial\hat{y}_i}{\partial t_j}=\begin{bmatrix}\frac{\partial\hat{y}_1}{\partial t_1}&\frac{\partial\hat{y}_1}{\partial t_2}&\frac{\partial\hat{y}_1}{\partial t_3}\\\frac{\partial\hat{y}_2}{\partial t_1}&\frac{\partial\hat{y}_2}{\partial t_2}&\frac{\partial\hat{y}_2}{\partial t_3}\\\frac{\partial\hat{y}_3}{\partial t_1}&\frac{\partial\hat{y}_3}{\partial t_2}&\frac{\partial\hat{y}_3}{\partial t_3}\end{bmatrix}$$

$$\frac{\partial t_i}{\partial z}=\frac{\partial}{\partial z}\left[w_iz+b_i\right]=w_i$$

$$\frac{\partial z}{\partial r_i}=\frac{1}{9}$$

$$\frac{\partial r_i}{\partial c_j}=\begin{cases}1&\text{if } c_j>0\\0&\text{otherwise}\end{cases}$$

$$\frac{\partial c_i}{\partial k_j}=\dots$$