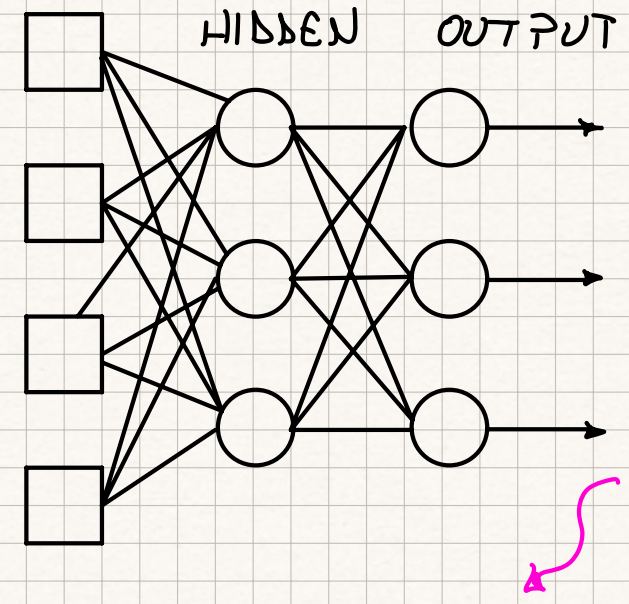
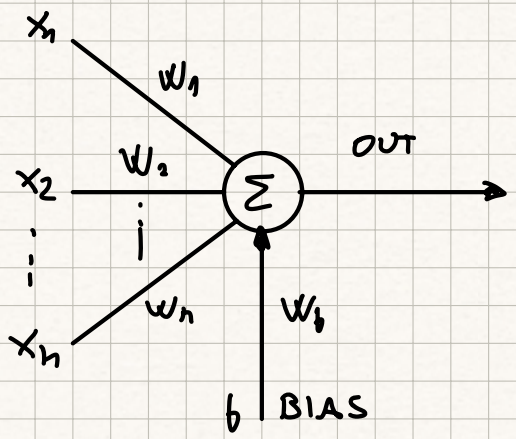
FORWARD PASS



MSE
ENTROPY
LOSS

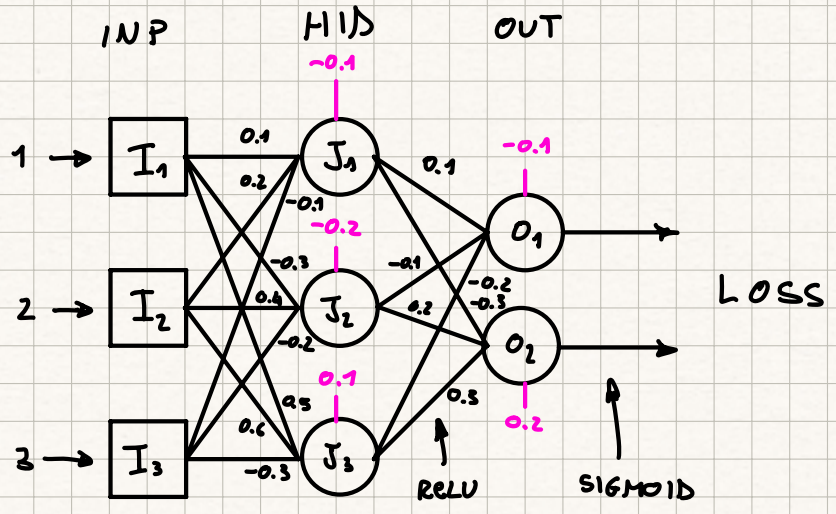


$$\Sigma = \left[\sum_{i=0} (x_i w_i)\right] + b w_b$$

$$OUT = f(\Sigma)$$

f = ReLU / sigmoid / Tanh ...

- INPUT ARE NOT ACTUALLY NEURON BUT ARE THE INPUT FROM THE IMAGE
- EACH INPUT I IS THE VALUE OF A SINGLE PIXEL FROM THE IMAGE
- ONE OUTPUT FOR EACH POSSIBLE CLASS OF THE OUTPUT
- $ReLU = max(0, in)$
- $Softmax = e^{net_x} / \sum_s e^{net_s}$ $\quad net = \Sigma$ $\hookrightarrow \in [0, 1]$



$J_1 = (0.1 + 0.4 - 0.3) + (-0.1) = 0.1 \rightarrow RELU(J_1) = 0.1$

$J_2 = (-0.3 + 0.8 - 0.6) + (-0.2) = -0.3 \rightarrow RELU(J_2) = 0$

$J_3 = (0.5 + 1.2 - 0.9) + 0.1 = 0.9 \rightarrow RELU(J_3) = 0.9$

$O_1 = (0.01 + 0 - 0.18) + (-0.1) = -0.27 \rightarrow SIG(O_1) = 0.432907$

$O_2 = (-0.03 + 0 + 0.27) + 0.2 = 0.44 \rightarrow SIG(O_2) = 0.608259$

$$MSE = \frac{1}{2}\left[(1 - 0.432907)^2 + (0 - 0.608259)^2\right] = 0.345787$$

$$MSE' = 2/2 \cdot (\hat{y} - y) = (-0.567093, 0.608259)$$

$$\frac{\partial L}{\partial J_{11}} = \frac{\partial L_1}{\partial J_{11}} = \frac{\partial L_1}{\partial O_1}\frac{\partial O_1}{\partial J_{11}} = \frac{\partial L_1}{\partial SIG_{OUT}}\frac{\partial SIG()}{\partial O_1}\cdot\frac{\partial O_1}{\partial J_{11}}$$

$$= (-0.567093)(0.432907(1 - 0.432907))(1 \cdot 0.1)$$

$$= (-0.567093)(0,24549853)(0.1) = -0.01392205$$

$$\frac{\partial L}{\partial J_{21}} = \frac{\partial L_1}{\partial J_{21}} = \frac{\partial L_1}{\partial O_1}\frac{\partial O_1}{\partial J_{21}} = \frac{\partial L_1}{\partial SIG_{OUT}}\frac{\partial SIG()}{\partial O_1}\cdot\frac{\partial O_1}{\partial J_{21}}$$

$$= (-0.567093)\left(0.432907(1-0.432907)\right)$$

$$= (-0.567093)(0,2454985\,3)\,0 = 0$$

$$\frac{\partial L}{\partial J_{31}} = \frac{\partial L_1}{\partial J_{31}} = \frac{\partial L_1}{\partial o_1}\,\frac{\partial o_1}{\partial J_{31}} = \frac{\partial L_1}{\partial \text{SIG}_{OUT}}\,\frac{\partial \text{SIG}()}{\partial o_1}\cdot\frac{\partial o_1}{\partial J_{31}}$$

$$= (-0.567093)\left(0.432907(1-0.432907)\right)(1\cdot 0.9)$$

$$= (-0.567093)(0,2454985\,3)\,0.9 = -0.1252985$$

$$\frac{\partial L}{\partial J_{12}} = \frac{\partial L_2}{\partial J_{12}} = \frac{\partial L_2}{\partial o_2}\,\frac{\partial o_2}{\partial J_{12}} = \frac{\partial L_2}{\partial \text{SIG}_{OUT}}\,\frac{\partial \text{SIG}()}{\partial o_2}\cdot\frac{\partial o_2}{\partial J_{12}}$$

$$= (0.608259)\left(0.608259(1-0.608259)\right)(1\cdot 0.1)$$

$$= (0.608259)(0.23827999)(0.1) = 0.01449359$$

$$\frac{\partial L}{\partial J_{22}} = \frac{\partial L_2}{\partial J_{22}} = \frac{\partial L_2}{\partial o_2}\,\frac{\partial o_2}{\partial J_{22}} = \frac{\partial L_2}{\partial \text{SIG}_{OUT}}\,\frac{\partial \text{SIG}()}{\partial o_2}\cdot\frac{\partial o_2}{\partial J_{22}}$$

$$= (0.608259)\left(0.608259(1-0.608259)\right)(1\cdot 0)$$

$$= (0.608259)(0.23827999)\,0 = 0$$

$$\frac{\partial L}{\partial J_{32}} = \frac{\partial L_2}{\partial J_{32}} = \frac{\partial L_2}{\partial o_2}\frac{\partial o_2}{\partial J_{32}} = \frac{\partial L_2}{\partial SIG_{OUT}}\frac{\partial SIG()}{\partial o_2}\cdot\frac{\partial o_2}{\partial J_{32}}$$

$$= (0.608259)(0.608259(1-0.608259))(1\cdot 0.9)$$

$$= (0.608259)(0.23827999)\,0.9 = 0.13044235$$

$$\partial B_{O_1} = \frac{\partial L_1}{\partial SIG_{OUT}}\frac{\partial SIG()}{\partial o_1} = -0,1392205$$

$$\partial B_{O_2} = \frac{\partial L_2}{\partial SIG_{OUT}}\frac{\partial SIG()}{\partial o_2} = 0.14493595$$

$$\frac{\partial L}{\partial I_{11}} = \frac{\partial L_1}{\partial I_{11}} + \frac{\partial L_2}{\partial I_{11}}$$

$$\rightarrow \frac{\partial L_1}{\partial I_{11}} = \frac{\partial L_1}{\partial SIG_{OUT}}\frac{\partial SIG()}{\partial o_1}\cdot\frac{\partial o_1}{\partial REL_{OUT}}\frac{\partial REL()}{\partial J_1}\frac{\partial J_1}{\partial I_{11}}$$

$$= (-0.567093)(0,2454853)\,0.1\cdot 1\cdot 1$$

(0.1 = WEIGHT J_{11})

$$\rightarrow \frac{\partial L}{\partial I_{11}} = \frac{\partial L_2}{\partial SIG_{OUT}}\frac{\partial SIG()}{\partial o_2}\cdot\frac{\partial o_2}{\partial REL_{OUT}}\frac{\partial REL()}{\partial J_1}\frac{\partial J_1}{\partial I_{11}}$$

$$= (0.608259)(0.23827999)(-0.3)\cdot 1\cdot 1$$

(-0.3 = WEIGHT J_{12})

$$\Rightarrow = (-0,013922) + (-0,043481)$$

$$= -0,0574028$$

$$\frac{\partial L}{\partial I_{21}} = \frac{\partial L_1}{\partial I_{21}} + \frac{\partial L_2}{\partial I_{21}}$$

$$\rightarrow \frac{\partial L_1}{\partial I_{21}} = \frac{\partial L_1}{\partial SIG_{OUT}} \frac{\partial SIG()}{\partial O_1} \cdot \frac{\partial O_1}{\partial REL_{OUT}} \frac{\partial REL()}{\partial J_1} \frac{\partial J_1}{\partial I_{21}}$$

$$= (-0.567093)(0,2454853)\, 0.1 \cdot 1 \cdot 2$$

0.1 ← WEIGHT J_{11}

$$\rightarrow \frac{\partial L}{\partial I_{21}} = \frac{\partial L_2}{\partial SIG_{OUT}} \frac{\partial SIG()}{\partial O_2} \cdot \frac{\partial O_2}{\partial REL_{OUT}} \frac{\partial REL()}{\partial J_1} \frac{\partial J_1}{\partial I_{21}}$$

$$= (0.608259)(0.2382799)(-0.3) \cdot 1 \cdot 2$$

-0.3 ← WEIGHT J_{12}

$$\Rightarrow = (-0.027844) + (0,086962)$$

$$= -0.114806$$

$$\frac{\partial L}{\partial I_{31}} = \frac{\partial L_1}{\partial I_{31}} + \frac{\partial L_2}{\partial I_{31}}$$

$$\rightarrow \frac{\partial L_1}{\partial I_{31}} = \frac{\partial L_1}{\partial SIG_{OUT}} \frac{\partial SIG()}{\partial O_1} \cdot \frac{\partial O_1}{\partial REL_{OUT}} \frac{\partial REL()}{\partial J_1} \frac{\partial J_1}{\partial I_{31}}$$

$$= (-0.567093)(0,2454853)\, 0.1 \cdot 1 \cdot 3$$

0.1 ← WEIGHT J_{11}

$$\rightarrow \frac{\partial L}{\partial I_{31}} = \frac{\partial L_2}{\partial SIG_{OUT}} \frac{\partial SIG()}{\partial O_2} \cdot \frac{\partial O_2}{\partial REL_{OUT}} \frac{\partial REL()}{\partial J_1} \frac{\partial J_1}{\partial I_{31}}$$

$$= (0.608259)(0.23827999)(-0.3) \cdot 1 \cdot 3$$

(-0.3) ← WEIGHT J_{12}

$$\Rightarrow = (-0,04176615) + (-0,13044235)$$

$$\hat{=} -0,1722085$$

$$\partial B_{J_1} = -0,057403$$

$$\Rightarrow \frac{\partial L}{\partial I_{12}} = \frac{\partial L}{\partial I_{22}} = \frac{\partial L}{\partial I_{32}} = \partial B_{J_2} = 0 \text{ since } \partial RELU\, J_2 = 0$$

$$\Rightarrow \frac{\partial L}{\partial I_{13}} = \frac{\partial L_1}{\partial I_{23}} + \frac{\partial L_2}{\partial I_{33}}$$

$$\rightarrow \frac{\partial L_1}{\partial I_{13}} = \frac{\partial L_1}{\partial SIG_{OUT}} \frac{\partial SIG()}{\partial O_1} \cdot \frac{\partial O_1}{\partial REL_{OUT}} \frac{\partial REL()}{\partial J_3} \frac{\partial J_1}{\partial I_{13}}$$

$$= (-0.567093)(0,24549853)(-0.2) \cdot 1 \cdot 1$$

(-0.2) ← WEIGHT J_{31}

$$\rightarrow \frac{\partial L}{\partial I_{13}} = \frac{\partial L_2}{\partial SIG_{OUT}} \frac{\partial SIG()}{\partial O_2} \cdot \frac{\partial O_2}{\partial REL_{OUT}} \frac{\partial REL()}{\partial J_3} \frac{\partial J_1}{\partial I_{13}}$$

$$= (0.608259)(0.23827999)(0.3) \cdot 1 \cdot 1$$

(0.3) ← WEIGHT J_{32}

$$\Rightarrow = (0,0278441) + (0,043480784 53)$$

$$\hat{=} 0,07132249 85$$

$$\Rightarrow \frac{\partial L}{\partial I_{23}} = \frac{\partial L_1}{\partial I_{23}} + \frac{\partial L_2}{\partial I_{23}}$$

$$\rightarrow \frac{\partial L_1}{\partial I_{23}} = \frac{\partial L_1}{\partial SIG_{OUT}} \frac{\partial SIG()}{\partial o_1} \cdot \frac{\partial o_1}{\partial REL_{OUT}} \frac{\partial REL()}{\partial J_3} \frac{\partial J_1}{\partial I_{23}}$$

$$= (-0.567093)(0,24549853)(0.2) \cdot 1 \cdot 2$$

WEIGHT J_{31}

$$\rightarrow \frac{\partial L}{\partial I_{23}} = \frac{\partial L_2}{\partial SIG_{OUT}} \frac{\partial SIG()}{\partial o_2} \cdot \frac{\partial o_2}{\partial REL_{OUT}} \frac{\partial REL()}{\partial J_3} \frac{\partial J_1}{\partial I_{23}}$$

$$= (0.608259)(0.23827999)(0.3) \cdot 1 \cdot 2$$

WEIGHT J_{32}

$$\Rightarrow = (0,0556882) + (0,0869157)$$

$$\stackrel{!}{=} 0,1426497$$

$$\Rightarrow \frac{\partial L}{\partial I_{33}} = \frac{\partial L_1}{\partial I_{33}} + \frac{\partial L_2}{\partial I_{33}}$$

$$\rightarrow \frac{\partial L_1}{\partial I_{33}} = \frac{\partial L_1}{\partial SIG_{OUT}} \frac{\partial SIG()}{\partial o_1} \cdot \frac{\partial o_1}{\partial REL_{OUT}} \frac{\partial REL()}{\partial J_3} \frac{\partial J_1}{\partial I_{33}}$$

$$= (-0.567093)(0,24549853)(0.2) \cdot 1 \cdot 3$$

WEIGHT J_{31}

$$\rightarrow \frac{\partial L}{\partial I_{33}} = \frac{\partial L_2}{\partial SIG_{OUT}} \frac{\partial SIG()}{\partial O_2} \cdot \frac{\partial O_2}{\partial REL_{OUT}} \frac{\partial REL()}{\partial J_3} \frac{\partial J_1}{\partial I_{33}}$$

$$= (0.608259)(0.23827999)(0.3) \cdot 1 \cdot 3$$

WEIGHT J_{32}

$$\Rightarrow = (0{,}0835323) + (0{,}13044235)$$

$$\doteq 0{,}213974654$$

$$\partial B_{J3} = \frac{\partial L}{\partial J_3} = 0{,}071322\,4885$$

DONE ✓

WEIGHTS/BIASES UPDATE

* SUPPOSE THIS IS THE AVERAGE VECTOR OF ONE BATCH AND WEIGHTS BIASES ARE SAME AS BEFORE

```
average[0].biases[0]: -0.028701
average[0].biases[1]: 0.000000
average[0].biases[2]: 0.071283
average[0].weights[0][0]: -0.028701
average[0].weights[0][1]: -0.057403
average[0].weights[0][2]: -0.086104
average[0].weights[1][0]: 0.000000
average[0].weights[1][1]: 0.000000
average[0].weights[1][2]: 0.000000
average[0].weights[2][0]: 0.106903
average[0].weights[2][1]: 0.106945
average[0].weights[2][2]: 0.213849
average[1].biases[0]: -0.139010
average[1].biases[1]: 0.144936
average[1].weights[0][0]: -0.006961
average[1].weights[0][1]: -0.000000
average[1].weights[0][2]: -0.118169
average[1].weights[1][0]: 0.007247
average[1].weights[1][1]: 0.000000
average[1].weights[1][2]: 0.123196
```

LAYER 1

LAYER 2

SUPPOSE LEARNING RATE 0.01

LAYER 1

$B1 : -0.1 - (0.01 \cdot A[0].b[0]) \approx -0,09971299$

$B2 : -0.2 - (0.01 \cdot A[0].b[1]) \approx -0.2$

$B3 : 0.1 - (0.01 \cdot A[0].b[2]) \approx 0,09928717$

$W_{11} : 0.1 - (0.01 \cdot A[0].W[0][0]) = 0,10028701$

$W_{12} : 0.2 - (0.01 \cdot A[0].W[0][1]) = 0,20057403$

$W_{13} : -0.1 - (0.01 \cdot A[0].W[0][2]) = -0,0999139$

$W_{21} : -0.3 - (0.01 \cdot A[0].W[1][0]) = -0,3$

$W_{22} : 0.4 - (0.01 \cdot A[0].W[1][1]) = 0,4$

$W_{23} : -0.2 - (0.01 \cdot A[0].W[1][2]) = -0,2$

$W_{31} : 0.5 - (0.01 \cdot A[0].W[2][0]) = 0,49893097$

$W_{32} : 0.6 - (0.01 \cdot A[0].W[2][1]) = 0,59893055$

$W_{33} : -0.3 - (0.01 \cdot A[0].W[2][2]) = -0,30213849$

LAYER 2

$B1 : -0.1 - (0.01 \cdot A[1].b[0]) \approx -0,0986099$

$B2 : 0.2 - (0.01 \cdot A[1].b[1]) \approx 0,19855062$

$W_{11} : 0.1 - (0.01 \cdot A[1].W[0][0]) = 0,10006961$

$W_{12} : -0.1 - (0.01 \cdot A[1].W[0][1]) = -0,1 \approx$

$W_{13} : -0.2 - (0.01 \cdot A[1].W[0][2]) = -0,19881831$

$W_{21} : -0.3 - (0.01 \cdot A[1].W[1][0]) = -0,30007247$

$W_{22} : 0.2 - (0.01 \cdot A[1].W[1][1]) = 0,2$

$W_{23} : 0.3 - (0.01 \cdot A[1].W[1][2]) = 0,29876804$

DONE V