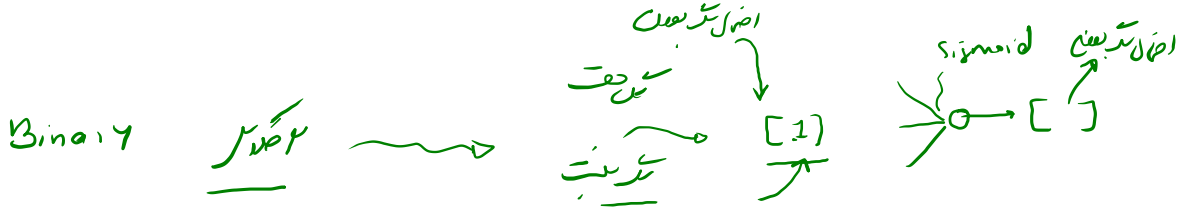
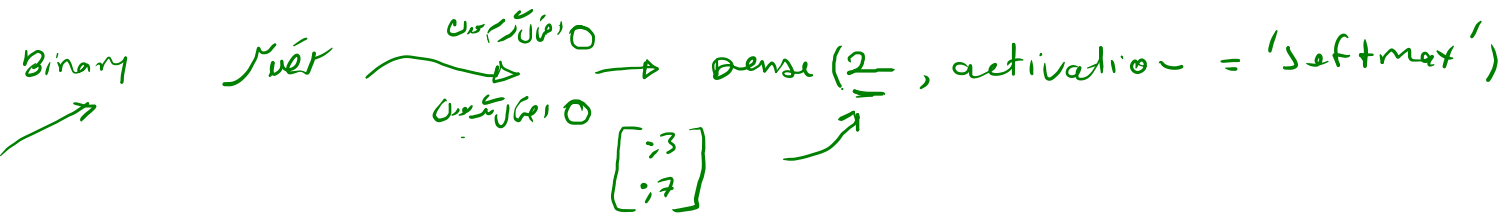
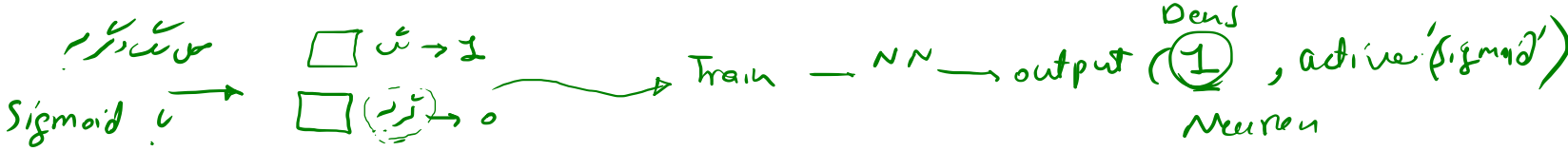
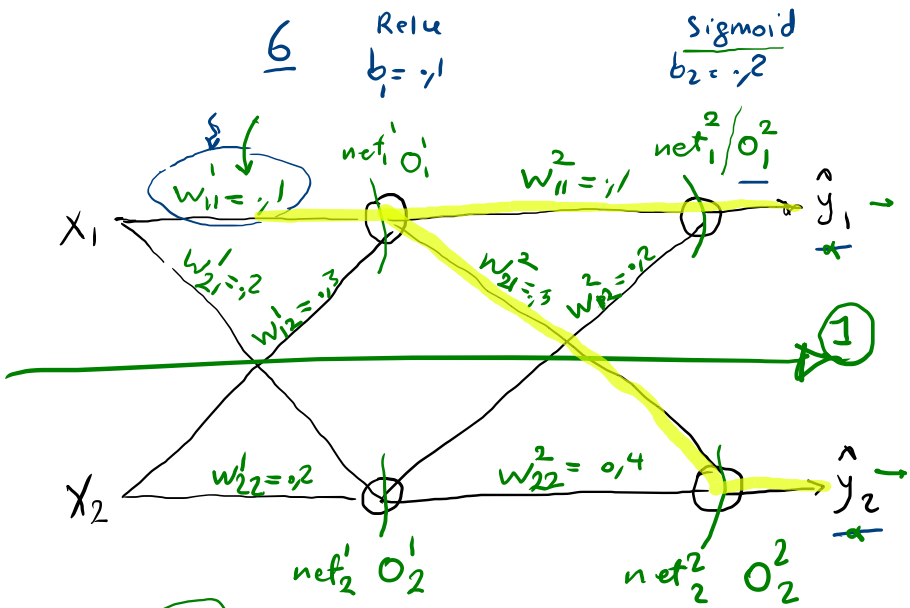
Back propagation

(Numerical Ex)


Binary
اصل تشخیص
[1]
Sigmoid
[ ]


Binary
اصل تشخیص
اصل تشخیص
Dense(2, activation = 'softmax')
[;3
;7]


Sigmoid
Train — NN — output
Dens
1
Neuron
, activation='sigmoid')



X_1	X_2
0,7	0,5

$y_1 True = ,3$

$y_2 True = ,7$

1.

$net_1^1 = w_{11}^1 X_1 + w_{12}^1 X_2 + b_1$

$O_1^1 = Relu(net_1^1)$

$net_2^1 = w_{21}^1 X_1 + w_{22}^1 X_2 + b_1$

$O_2^1 = Relu(net_2^1)$

$net_1^2 = w_{11}^2 O_1^1 + w_{12}^2 O_2^1 + b_2$

$O_1^2 = 6(net_1^2) \rightarrow \hat{y}_1$

$net_2^2 = w_{21}^2 O_1^1 + w_{22}^2 O_2^1 + b_2$

$O_2^2 = 6(net_2^2) \rightarrow \hat{y}_2$

② $L = MSE = \frac{1}{N} \sum_{i=1}^{N} (y_{i_{true}} - \hat{y}_i)^2$

SGD ↓

به ازای هر sample که داریم

یک گرادیان محاسبه، بشود و وزن‌ها

آپدیت بشود!

↓ SGD

$L = (y_{true} - \hat{y})^2$

$\circ\ \hat{y}_1$

$\circ\ \hat{y}_2$

$\begin{cases} \hat{y}_1 \\ \hat{y}_2 \end{cases}$

$L = \frac{1}{2}\left[(0{,}3 - \hat{y}_1)^2 + (0{,}7 - \hat{y}_2)^2 \right]$

$\frac{\partial L}{\partial \hat{y}_2} = \frac{-1}{\cancel{2}}(\cancel{2})(0{,}7 - \hat{y}_2)$

$\frac{\partial L}{\partial \hat{y}_1} = \frac{-1}{\cancel{2}}(\cancel{2})(0{,}3 - \hat{y}_1)$

③ Back propagation

target : w_{11}^1

$$w_{11}^1\text{-new} = w_{11}^1\text{-old} - \alpha \frac{\partial L}{\partial w_{11}^1}$$

0/01

$$\frac{\partial L}{\partial w_{11}^1} = \left[\frac{\partial L}{\partial \hat{y}_1}\frac{\partial \hat{y}_1}{\partial o_1^2}\frac{\partial o_1^2}{\partial net_1^2}\frac{\partial net_1^2}{\partial o_1^1}\right]\left[\frac{\partial o_1^1}{\partial net_1^1}\frac{\partial net_1^1}{\partial w_{11}^1}\right] + \left[\frac{\partial L}{\partial \hat{y}_2}\frac{\partial \hat{y}_2}{\partial o_2^2}\frac{\partial o_2^2}{\partial net_2^2}\frac{\partial net_2^2}{\partial o_1^1}\right]\left[\frac{\partial o_1^1}{\partial net_1^1}\frac{\partial net_1^1}{\partial w_{11}^1}\right]$$

$$\frac{\partial L}{\partial w_{11}^1} = \left[\frac{\partial L}{\partial \hat{y}_1}\frac{\partial \hat{y}_1}{\partial o_1^2}\frac{\partial o_1^2}{\partial net_1^2}\frac{\partial net_1^2}{\partial o_1^1} + \frac{\partial L}{\partial \hat{y}_2}\frac{\partial \hat{y}_2}{\partial o_2^2}\frac{\partial o_2^2}{\partial net_2^2}\frac{\partial net_2^2}{\partial o_1^1}\right]\left[\frac{\partial o_1^1}{\partial net_1^1}\frac{\partial net_1^1}{\partial w_{11}^1}\right]$$

$\frac{\partial L}{\partial \hat{y}_1}$: $(\hat{y}_1 - y_1)$; $\frac{\partial \hat{y}_1}{\partial o_1^2}$: 1 ; $\frac{\partial o_1^2}{\partial net_1^2}$: $\frac{\partial \sigma(net_1^2)}{\partial net_1^2}$; $\frac{\partial net_1^2}{\partial o_1^1}$: w_{11}^2

$\frac{\partial L}{\partial \hat{y}_2}$: $(y_2 - \hat{y}_2)$; $\frac{\partial \hat{y}_2}{\partial o_2^2}$: 1 ; $\frac{\partial o_2^2}{\partial net_2^2}$: $\frac{\partial \sigma(net_2^2)}{\partial net_2^2}$; $\frac{\partial net_2^2}{\partial o_1^1}$: w_{21}^2

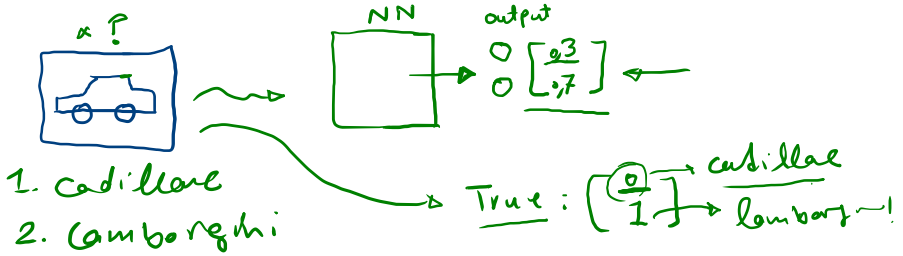
$\frac{\partial o_1^1}{\partial net_1^1}$: $\frac{\partial Relu(net_1^1)}{\partial net_1^1}$; $\frac{\partial net_1^1}{\partial w_{11}^1}$: x_1

MSE ⇝ loss function ⇝ regression!

cross-entropy ⇝ classification!

$$\rightsquigarrow l = -\sum_{i}^{N} y_i \cdot \log \hat{y}_{i\,pred}$$

N: Number of categ.



$$loss = -[y_1 \log y_{1\,pred} + y_2 \log y_{2\,pred}]$$

$$loss = -[0 \log 0.3 + 1 \log(0.7)] = -\log(0.7) =$$

$$-(-0.15) = 0.15$$

Cadillac

Lamborgini

NN

$\begin{bmatrix} .4 \\ .6 \end{bmatrix}$ ①

$\begin{bmatrix} .1 \\ .9 \end{bmatrix}$ ②

true

$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

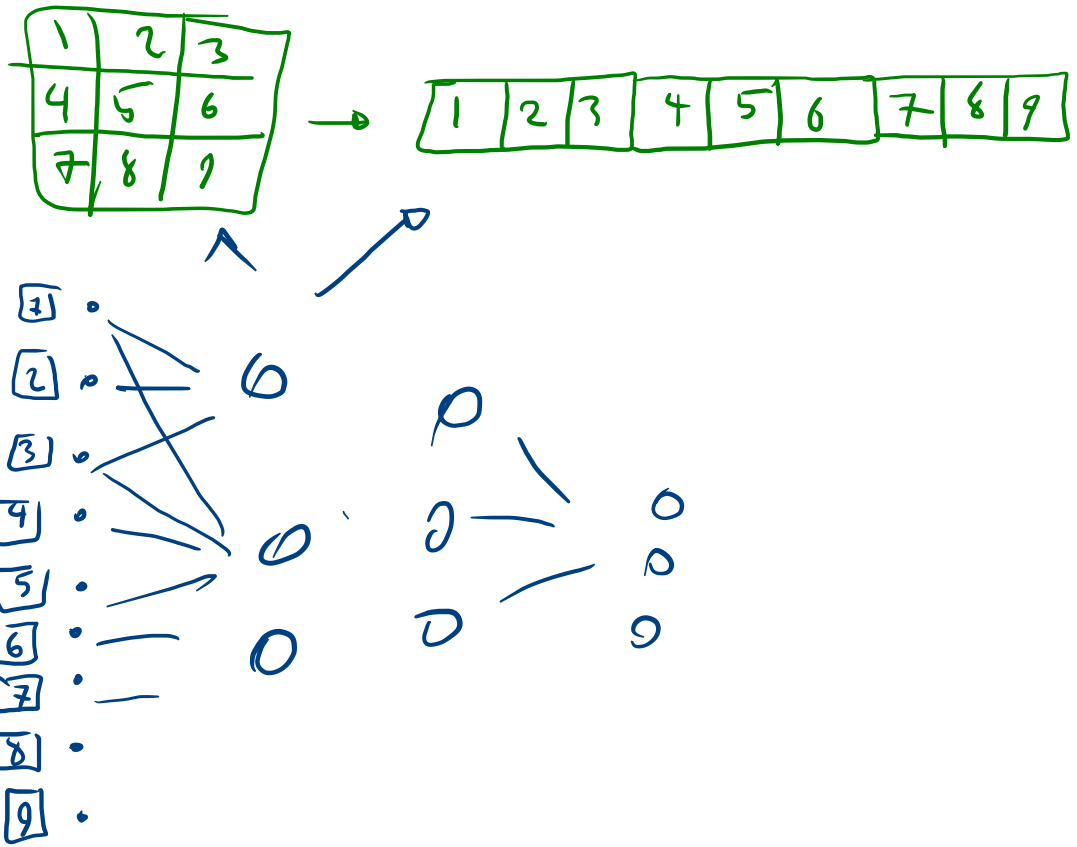
$\text{cross-entropy-1} = -[1 \log .6] = -(-.22) = .22$

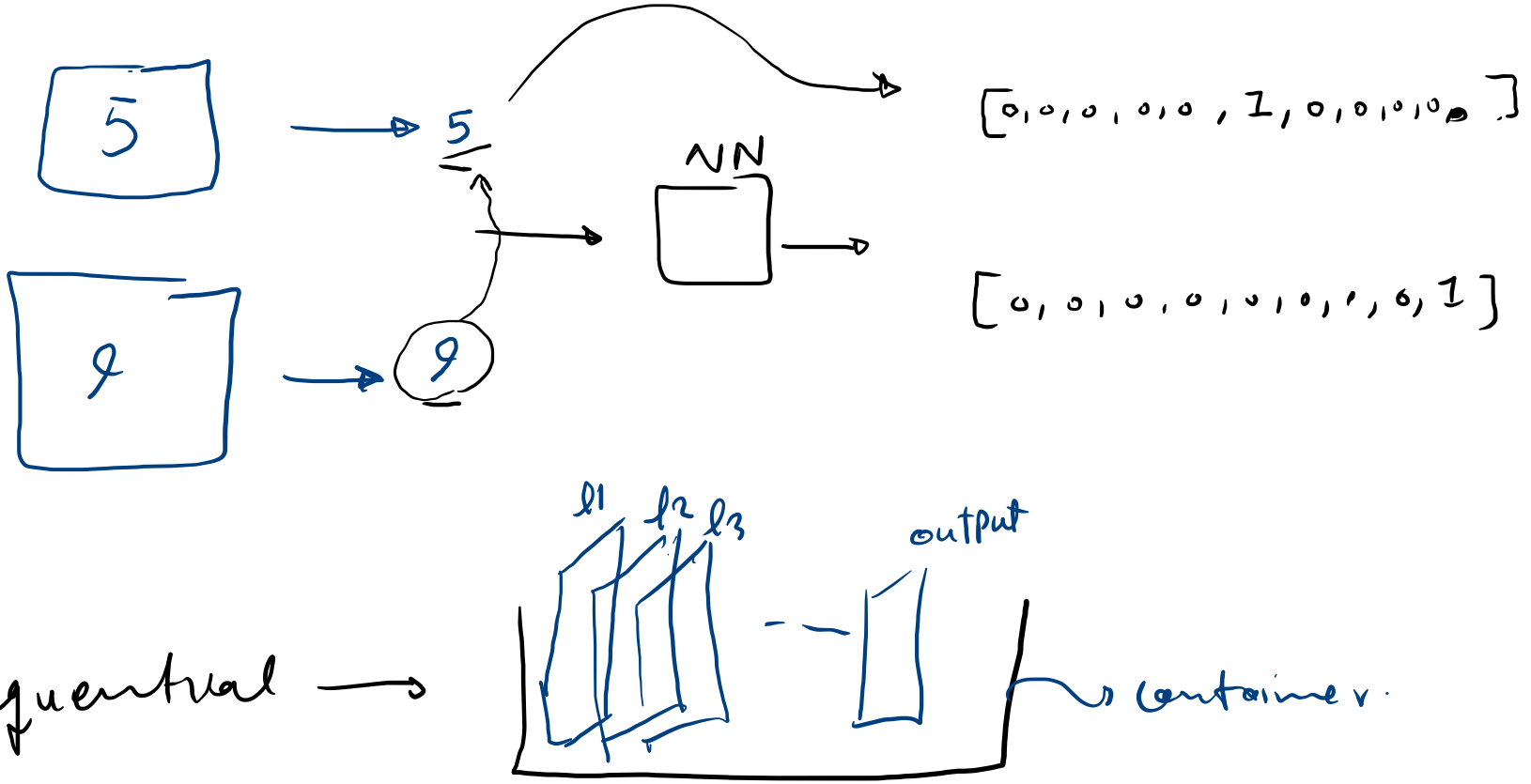
$\text{cross-entropy-2} = -[1 \log .9] = -(-0.045) = 0.045$ ←

Number of Category ← (N)

$$\text{cross} = -\sum_{i=1}^{N} y_{i\,\text{true}} \log y_{i\,\text{pred}}$$

Dense → Image !




5
5
NN
9
9
[0,0,0,0,0,1,0,0,0,0]
[0,0,0,0,0,0,0,0,0,1]
l1
l2
l3
output
Sequential
container

Relu $\begin{cases} \frac{\partial x}{\partial x} = 1 \end{cases}$

$$\frac{\partial \text{Relu}(x)}{\partial x} = ?$$

$$\text{Relu}(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$\frac{\partial \text{Relu}(x)}{\partial x} = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$$

$x = 0 \rightarrow$ undefined

$$\frac{\partial \text{Relu}(net_1^1)}{\partial net_1^1} = \begin{cases} 1 & net_1^1 > 0 \\ 0 & net_1^1 < 0 \end{cases}$$

$$\frac{\partial \sigma(x)}{\partial x} = \sigma(x)(1 - \sigma(x))$$

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$x = net_2^2 \Rightarrow \sigma(net_2^2)(1 - \sigma(net_2^2))$

$net_2^2 = 0.2$

$= \sigma(0.2)(1 - \sigma(0.2))$