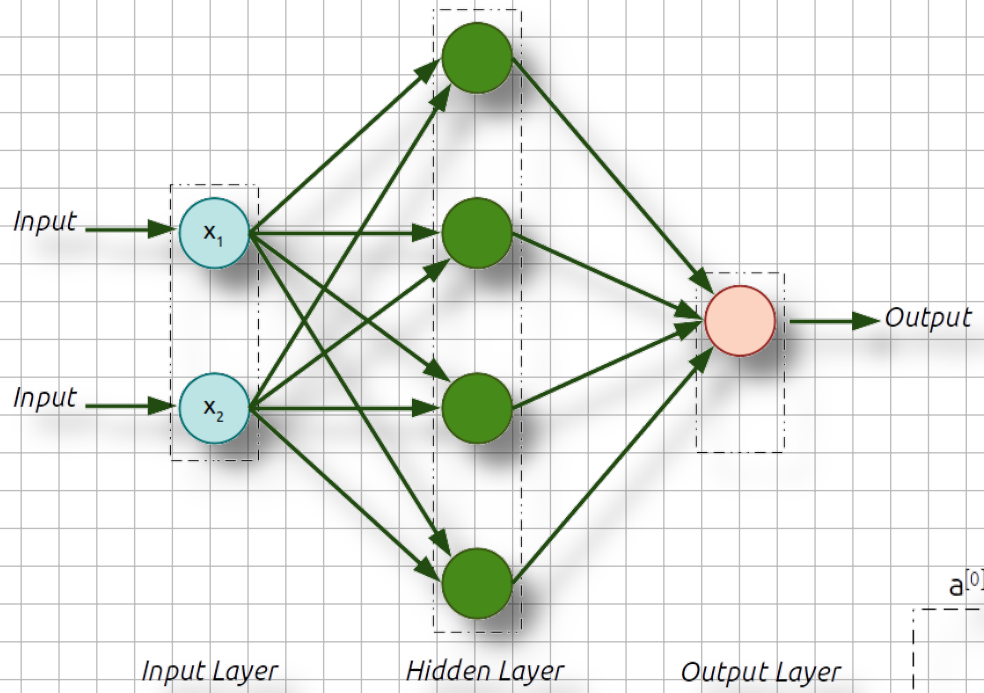
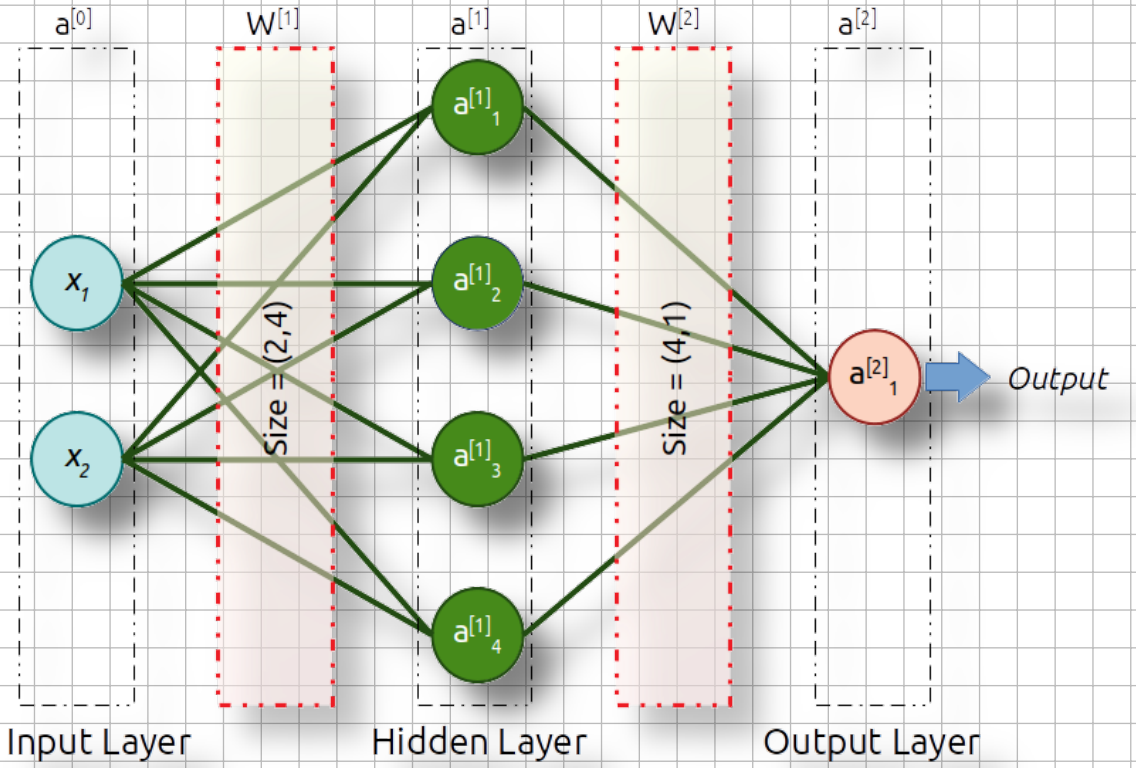
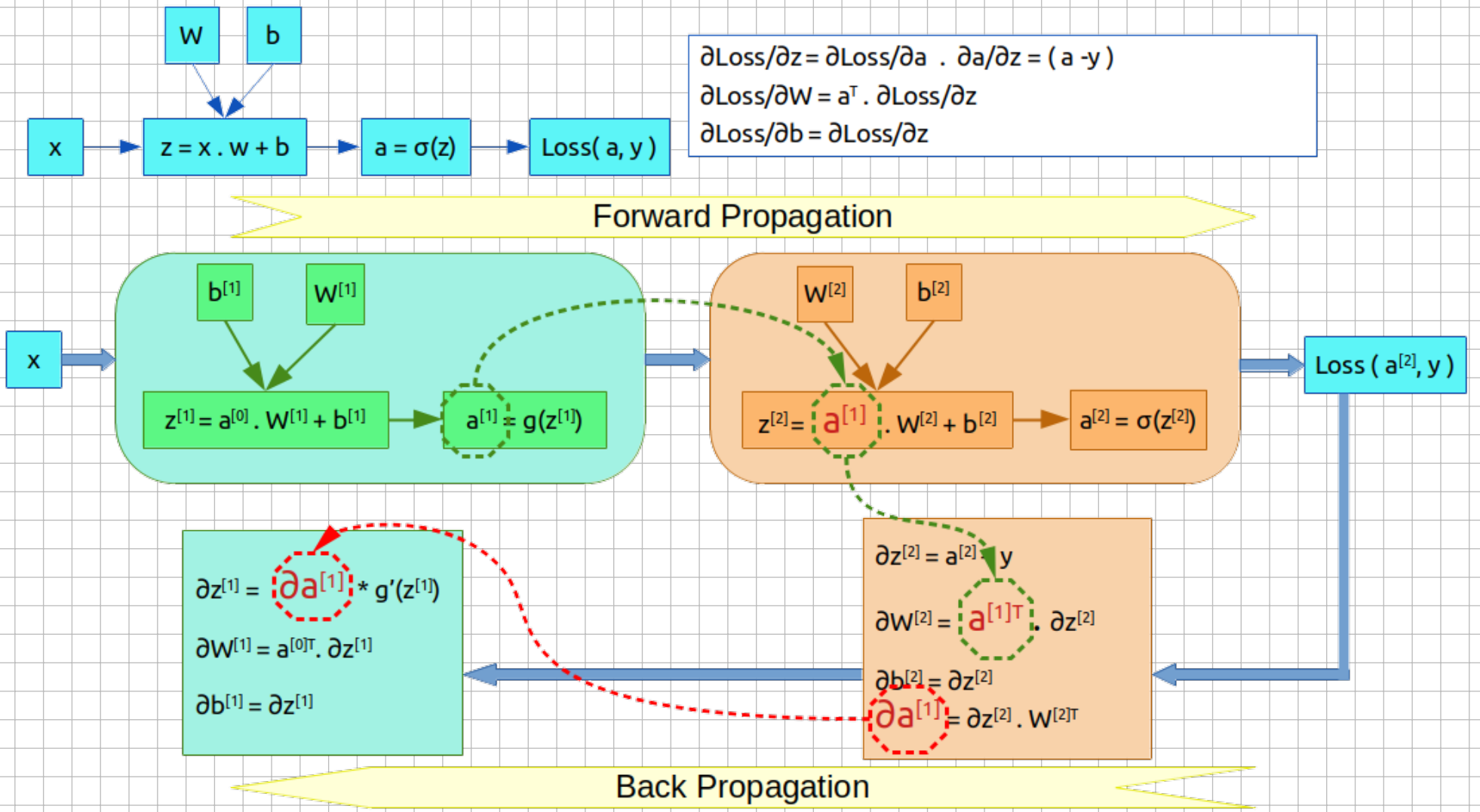

Input
Input
x_1
x_2
Output
Input Layer
Hidden Layer
Output Layer


$a^{[0]}$
$W^{[1]}$
$a^{[1]}$
$W^{[2]}$
$a^{[2]}$
x_1
x_2
$a^{[1]}_1$
$a^{[1]}_2$
$a^{[1]}_3$
$a^{[1]}_4$
$a^{[2]}_1$
Size = (2,4)
Size = (4,1)
Output
Input Layer
Hidden Layer
Output Layer



$\partial Loss/\partial z^{[1]} = \partial Loss/\partial z^{[2]} . \partial z^{[2]}/\partial a^{[1]} . \partial a^{[1]}/\partial z^{[1]} = \partial Loss/\partial a^{[1]} . g'(z^{[1]})$

$\partial Loss/\partial w^{[1]} = \partial Loss/\partial z^{[1]} . \partial z^{[1]}/\partial w^{[1]} = a^{[0]T} . \partial Loss/\partial z^{[1]}$

$\partial Loss/\partial a^{[0]} = \partial Loss/\partial z^{[1]} . \partial z^{[1]}/\partial a^{[0]} = \partial Loss/\partial z^{[1]} \cdot w^{[1]T}$

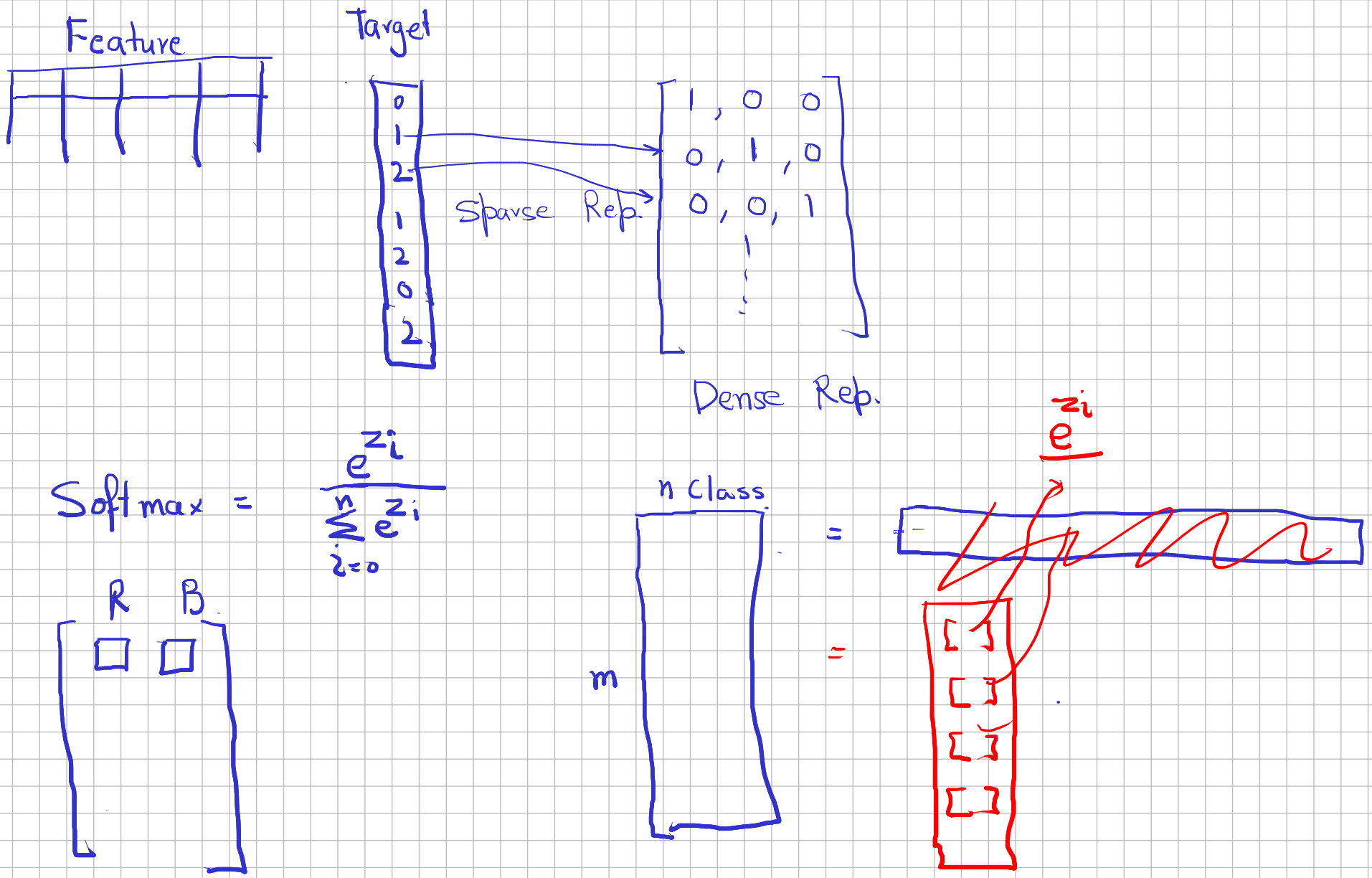
$\partial Loss/\partial b^{[1]} = \partial Loss/\partial z^{[1]}$

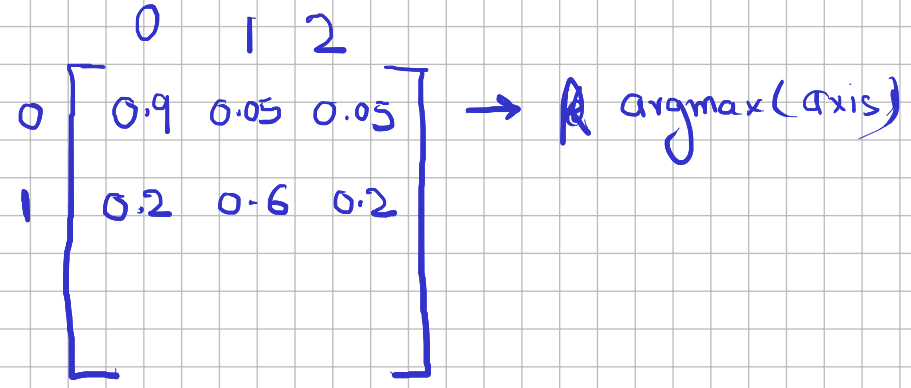
$\partial Loss/\partial z^{[2]} = \partial Loss/\partial a^{[2]} . \partial a^{[2]}/\partial z^{[2]} = (a^{[2]} - y)$

$\partial Loss/\partial w^{[2]} = \partial Loss/\partial z^{[2]} . \partial z^{[2]}/\partial w^{[2]} = a^{[1]T} \cdot \partial Loss/\partial z^{[2]}$

$\partial Loss/\partial a^{[1]} = \partial Loss/\partial z^{[2]} . \partial z^{[2]}/\partial a^{[1]} = \partial Loss/\partial z^{[2]} \cdot w^{[2]T}$

$\partial Loss/\partial b^{[2]} = \partial Loss/\partial z^{[2]} . \partial z^{[2]}/\partial b^{[2]} = \partial Loss/\partial z^{[2]}$


Feature
Target
0
1
2
1
2
0
2
Sparse Rep.
1, 0, 0
0, 1, 0
0, 0, 1
Dense Rep.
Softmax =
$\frac{e^{z_i}}{\sum_{i=0}^{n} e^{z_i}}$
R
B
n Class
m
=
=
e^{z_i}


0 1 2
0
1
0.9 0.05 0.05
0.2 0.6 0.2
argmax(axis)