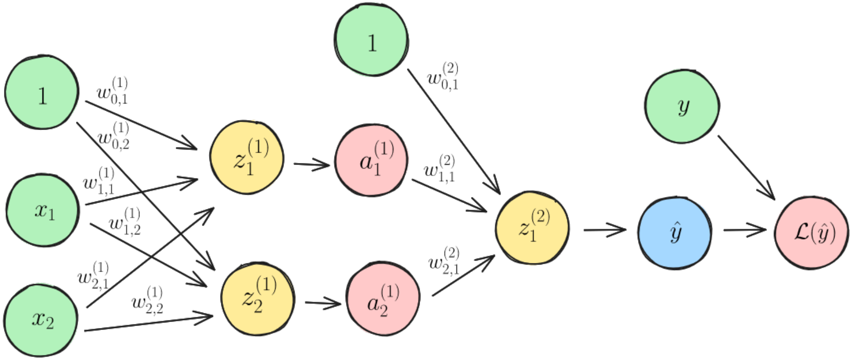
MLP exercise

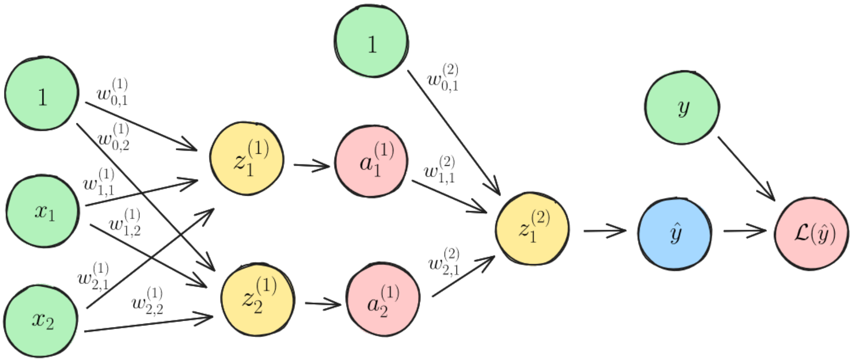


Let's calculate the MSE $\mathcal{L}(\hat{y}) = \frac{1}{2}(y - \hat{y})^2$, with $x_1 = 2$ and $x_2 = 3$, and a target $y = 10$, given that:

$w_{0,1}^{(1)} = -1$, $w_{1,1}^{(1)} = 2$, $w_{2,1}^{(1)} = 1$, $w_{0,2}^{(1)} = 2$, $w_{1,2}^{(1)} = -1$, $w_{2,2}^{(1)} = -1$,

$w_{0,1}^{(2)} = 1$, $w_{1,1}^{(2)} = 2$, $w_{2,1}^{(2)} = -1$

MLP exercise

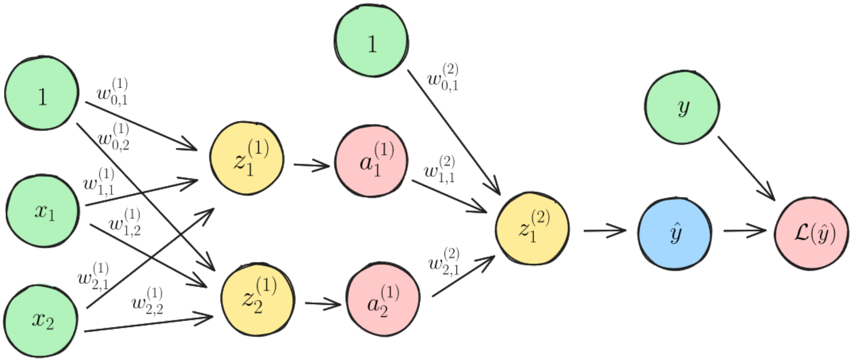


$a_1^{(1)} = 6$, $a_2^{(1)} = 0$

$\hat{y} = 13$,

$\mathcal{L}(\hat{y}) = 4.5$

MLP exercise

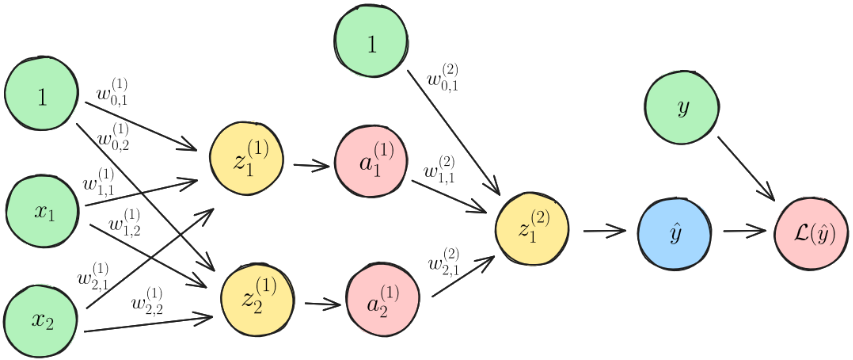


$a_1^{(1)} = 6$, $a_2^{(1)} = 0$

$\hat{y} = 13$,

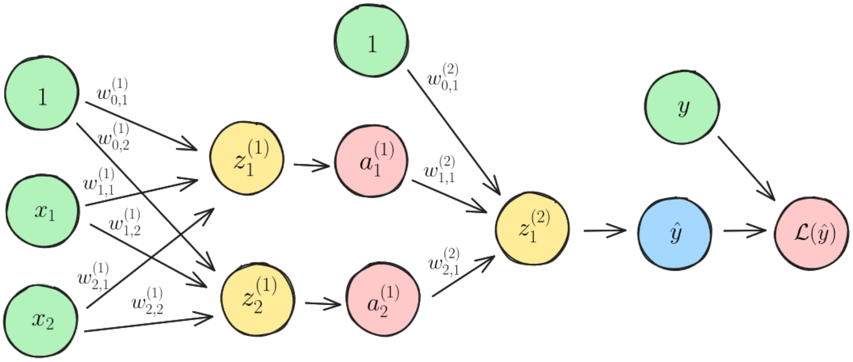
$\mathcal{L}(\hat{y}) = 4.5$ Now, let's compute the SGD update for weight $w_{2,1}^{(1)}$

MLP exercise



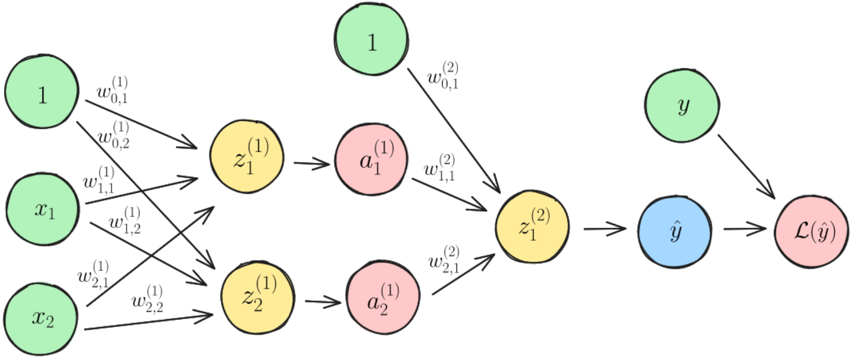
Now, let's compute the SGD update for weight $w_{2,1}^{(1)}$

MLP exercise



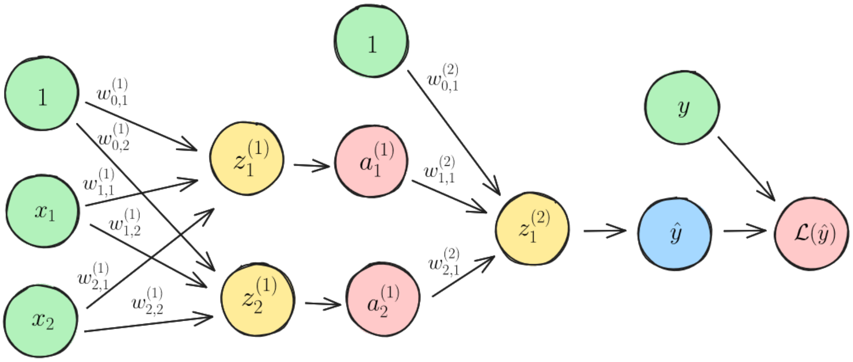
$$\Delta w_{2,1}^{(1)} = -\eta \frac{\partial \mathcal{L}}{\partial w_{2,1}^{(1)}}$$

MLP exercise



$$\Delta w_{2,1}^{(1)} = -\eta \frac{\partial \mathcal{L}}{\partial w_{2,1}^{(1)}} = -\eta \frac{\partial \mathcal{L}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_1^{(2)}} \cdot \frac{\partial z_1^{(2)}}{\partial a_1^{(1)}} \cdot \frac{\partial a_1^{(1)}}{\partial z_1^{(1)}} \cdot \frac{\partial z_1^{(1)}}{\partial w_{2,1}^{(1)}}$$

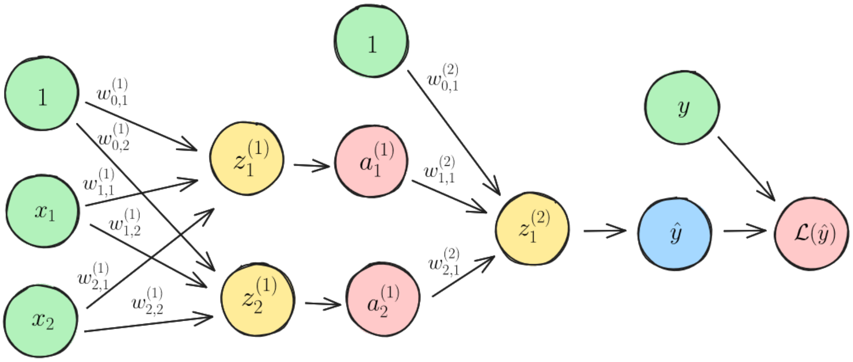
MLP exercise



$$\Delta w^{(1)}_{2,1} = -\eta \frac{\partial \mathcal{L}}{\partial w^{(1)}_{2,1}} = -\eta \frac{\partial \mathcal{L}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z^{(2)}_1} \cdot \frac{\partial z^{(2)}_1}{\partial a^{(1)}_1} \cdot \frac{\partial a^{(1)}_1}{\partial z^{(1)}_1} \cdot \frac{\partial z^{(1)}_1}{\partial w^{(1)}_{2,1}}$$

$$= -\eta (\hat{y} - y) \cdot 1 \cdot w^{(2)}_{1,1} \cdot 1 \cdot x_2$$

MLP exercise

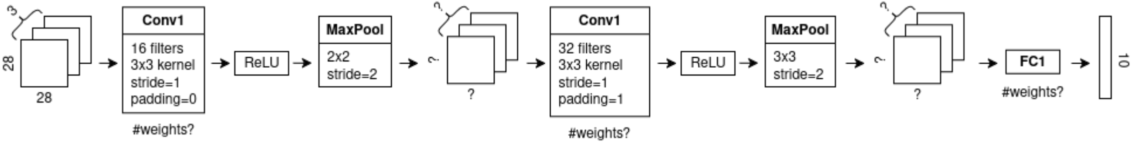


$$\Delta w_{2,1}^{(1)} = -\eta \frac{\partial \mathcal{L}}{\partial w_{2,1}^{(1)}} = -\eta \frac{\partial \mathcal{L}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_1^{(2)}} \cdot \frac{\partial z_1^{(2)}}{\partial a_1^{(1)}} \cdot \frac{\partial a_1^{(1)}}{\partial z_1^{(1)}} \cdot \frac{\partial z_1^{(1)}}{\partial w_{2,1}^{(1)}}$$

$$= -\eta(\hat{y} - y) \cdot 1 \cdot w_{1,1}^{(2)} \cdot 1 \cdot x_2 = -\eta(13 - 10) \cdot 2 \cdot 3 = -18\eta$$

We need to decrease $w_{2,1}^{(1)}$ to decrease the loss

CNN exercise



1. What are the dimensions of the intermediate feature maps?
2. How many weights are in the network?

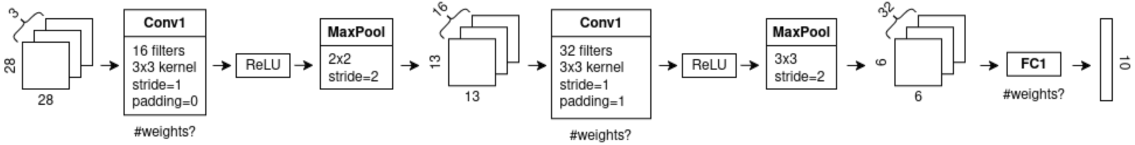
Reminder: convolutional layer

- Input size (of the layer and of every neuron): Channel $\times$ Width $\times$ Height
- Output size (of a neuron):

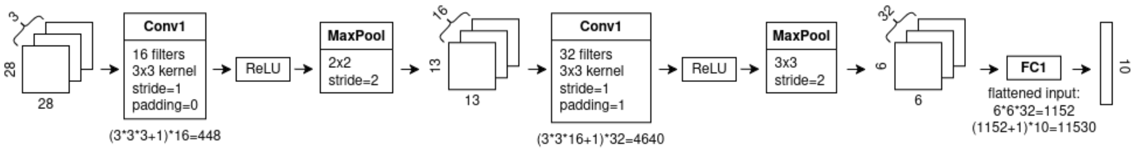
$$\underbrace{W_{out}}_{\text{width or height}} = \Bigg\lfloor \underbrace{\frac{\overbrace{W_{in} + 2 \times padding}^{\text{Input total width}} - kernel\_size}{stride}}_{\text{Number of possible kernel positions}} \Bigg\rfloor \underbrace{+1}_{\text{Starting position}}$$

- Output size (of a layer): Number of neurons $\times$ W_{out} $\times$ H_{out}

Feature maps dimensions



Number of weights



Total: 16618