

Exercise 5

Helin Marden

1.

2)

Early Stopping applies when the validation loss stops improving (doesn't decrease). It's purpose is to prevent the model from overfitting. It applies even when the training loss decreases. it solely depends on the validation loss.

b)

We look several steps ahead so we can make sure that the validation loss stagnation isn't a fluctuation but instead also applies on the coming steps.

2.
 2) Sample 1 mask (0, 1, 1, 0):

$$\tilde{x}^1 = (0, 1)$$

$$z_{in}^1 = W^T \tilde{x}^1 + b = \begin{pmatrix} -1 & 3 \\ 2 & -4 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ -4 \end{pmatrix} + \begin{pmatrix} 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

$$\text{ReLU}(z_{in}^1) = (6, 0) = z_{out}^1$$

keep $z_{1,out}^1$, drop $z_{2,out}^1 \rightarrow h^1 = (6, 0)$

$$f_{in}^1 = v^T \cdot h^1 + c = (1 \ 2) \cdot \begin{pmatrix} 6 \\ 0 \end{pmatrix} + 1 = 6 + 1 = 7$$

$$f_{out}^1 = \text{ReLU}(7) = 7 = \hat{y}^1$$

$$\delta_{out}^1 = \hat{y}^1 - y^1 = 7 - 2 = 5$$

$$\delta_h^1 = \frac{dL}{dh^1} = \frac{dL}{d\hat{y}^1} \cdot \frac{d\hat{y}^1}{dz_{in}^1} \cdot \frac{dz_{in}^1}{dh^1}$$

$$\delta_h^1 = (\hat{y}^1 - y^1) \cdot \text{ReLU}'(z_{in}^1) \cdot v = 5 \cdot 1 \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 10 \end{pmatrix}$$

$$\delta_{z_{in}}^1 = \begin{pmatrix} 5 \\ 10 \end{pmatrix} \cdot \text{ReLU}'(z_{in}^1) \cdot \text{Dropout} = (5, 10) \odot (1, 0) \odot (1, 0) = (5, 0)$$

GD:

$$\frac{dL}{dw}^1 = x \cdot \delta_{z_{in}}^T = \begin{pmatrix} 0 \\ 1 \end{pmatrix} (5 \ 0) = \begin{pmatrix} 0 & 0 \\ 5 & 0 \end{pmatrix}$$

$$\frac{dL}{db}^1 = \delta_{z_{in}} = \begin{pmatrix} 5 \\ 0 \end{pmatrix}$$

$$\frac{dL}{dv}^1 = \delta_{out} \cdot h = 5 \cdot \begin{pmatrix} 6 \\ 0 \end{pmatrix} = \begin{pmatrix} 30 \\ 0 \end{pmatrix}$$

$$\frac{dL}{dc}^1 = \delta_{out} = 5$$

Sample 2 mode (1, 1, 0, 1):

$$\tilde{x}^2 = (2, 2)$$

$$z_{in}^2 = W^T \tilde{x}^2 + b = \begin{pmatrix} -1 & 3 \\ 2 & -4 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 4 \\ -4 \end{pmatrix} + \begin{pmatrix} 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 7 \\ -1 \end{pmatrix}$$

$$z_{out}^2 = \text{ReLU}(7, -1) = (7, 0)$$

$$h^2 = (0, 0)$$

$$f_{in}^2 = v^T \cdot h^2 + c = (1, 2) \cdot \begin{pmatrix} 0 \\ 0 \end{pmatrix} + 1 = 0 + 1 = 1$$

$$\hat{y}^2 = f_{out}^2 = \text{ReLU}(1) = 1$$

$$\delta_{out}^2 = 1 - 3 = -2$$

$$\delta_h^2 = -2 \cdot \text{ReLU}'(1) \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 \\ -4 \end{pmatrix}$$

$$\delta_{z_{in}}^2 = \begin{pmatrix} -2 \\ -4 \end{pmatrix} \cdot \text{ReLU}'(7, -1) \cdot \text{Dropout} = \begin{pmatrix} -2 \\ -4 \end{pmatrix} \odot (1, 0) \odot (0, 1) = (0, 0)$$

$$\frac{dL^2}{dW} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \frac{dL^2}{db} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \frac{dL^2}{dv} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \frac{dL^2}{dc} = -2$$

$$W_{new} = W - \alpha \cdot \frac{dL^1}{dW} + \frac{dL^2}{dW} = \begin{pmatrix} -1 & 2 \\ 3 & -4 \end{pmatrix} - 0,1 \cdot \frac{\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}}{2} - \begin{pmatrix} -1 & 2 \\ 3 & -4 \end{pmatrix} - 0,1 \cdot \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 2 \\ 3 & -4 \end{pmatrix}$$

$$b_{new} = b - \alpha \cdot \frac{dL^1}{db} + \frac{dL^2}{db} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} - 0,1 \cdot \frac{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}{2} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} - 0,1 \cdot \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$$

$$v_{new} = v - \alpha \cdot \frac{\nabla L^2}{\nabla v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} - 0,1 \cdot \frac{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}{2} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} - 0,1 \cdot \begin{pmatrix} 15 \\ 0 \end{pmatrix} = \begin{pmatrix} 0,5 \\ 2 \end{pmatrix}$$

$$c_{new} = c - \alpha \cdot \frac{\nabla L^2}{\nabla c} = 1 - 0,1 \cdot \frac{5-2}{2} = 0,85$$

$$b) \quad x = (1, 2)$$

$$z_{in} = w^T x + b = \begin{pmatrix} -1 & 2,75 \\ 2 & -4 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 2,75 \\ 3 \end{pmatrix} = \begin{pmatrix} 4,5 \\ -6 \end{pmatrix} + \begin{pmatrix} 2,75 \\ 3 \end{pmatrix} = \begin{pmatrix} 7,25 \\ -3 \end{pmatrix}$$

$$z_{out} = \text{ReLU}(7,25, -1) = (5, 0)$$

$$f_{in} = v^T \cdot z_{out} + c = (-0,5 \ 2) \begin{pmatrix} 5 \\ 0 \end{pmatrix} + 0,85 = -1,65$$

$$\hat{y} = f_{out} = \text{ReLU}(-1,65) = 0$$