Project 1 overview

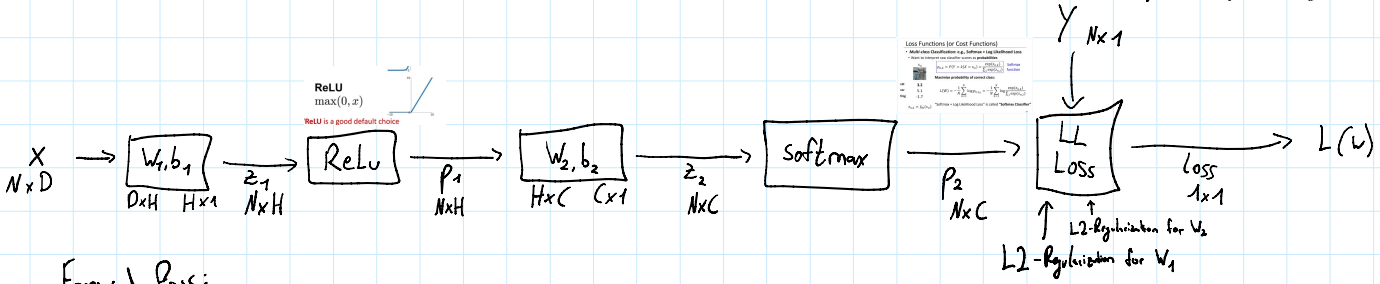
Samstag, 15. Oktober 2022 03:50

input - fully connected layer - ReLU - fully connected layer - softmax

W1: First layer weights; has shape (D, H)
b1: First layer biases; has shape (H,)
W2: Second layer weights; has shape (H, C)
b2: Second layer biases; has shape (C,)
Inputs:
- input_size: The dimension D of the input data.
- hidden_size: The number of neurons H in the hidden layer.
- output_size: The number of classes C.

$\longrightarrow X \in \mathbb{R}^{N \times D}$

$y[i] \in \{0, \dots, C-1\}$



Forward Pass:

$$z_1 = \underset{N\times D}{X} \cdot \underset{D\times H}{W_1} + \underset{N\times 1}{\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}} \cdot \underset{1\times H}{b_1^T} \in \mathbb{R}^{N\times H}$$

$$p_1 = \max(0, \underset{N\times H}{z_1}) \in \mathbb{R}^{N\times H}$$

$$z_2 = \underset{N\times H}{p_1} \cdot \underset{H\times C}{W_2} + \underset{N\times 1}{\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}} \cdot \underset{1\times C}{b_2^T} \in \mathbb{R}^{N\times C}$$

$$p_2[i][c] = \frac{\exp(z_2[i,c])}{\sum_{j=1}^{C} \exp(z_2[i,j])} \quad c \in \{1, \dots, C\} \qquad \left| \quad p_2 = \frac{\exp(z_2)}{\text{sum}(\exp(z_2), axis=1, keepdims=true)} \in \mathbb{R}^{N\times C} \right.$$

$$L(W) = -\frac{1}{N} \cdot \sum_{i=1}^{N} \log(p_2[i, y[i]]) + L2(W_1) + L2(W_2) \in \mathbb{R}^{1\times 1}$$

with $L2(W_1) = reg \cdot \text{sum}(W_1 ** 2)$, $L2(W_2) = reg \cdot \text{sum}(W_2 ** 2)$

Backpropagation, Gradients: (biases still have to be adjusted to batch-average)

$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial z_2} = \frac{p_2}{\partial z_2} \cdot \frac{\partial L}{\partial p_2}$$

$$= \underset{N\times C}{p_2} - \underset{N\times C}{(1 \text{ where } [i, y[i]])} \in \mathbb{R}^{N\times C} \quad (\text{average over } N)$$

$$\frac{\partial L}{\partial W_2} = \left(\frac{\partial L}{\partial W_{2,1}}, \frac{\partial L}{\partial W_{2,2}}, \dots, \frac{\partial L}{\partial W_{2,C}}\right)^T = \left(\frac{\partial z_2}{\partial W_{2,1}} \cdot \frac{\partial L}{\partial z_2}, \frac{\partial z_2}{\partial W_{2,2}} \cdot \frac{\partial L}{\partial z_2}, \dots, \frac{\partial z_2}{\partial W_{2,C}} \cdot \frac{\partial L}{\partial z_2}\right)^T$$

$$= \underset{H\times N}{p_1^T} \cdot \underset{N\times C}{\frac{\partial L}{\partial z_2}} \in \mathbb{R}^{H\times C}$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial L}{\partial z_1} = \frac{\partial L}{\partial z_1} = \frac{\partial p_1}{\partial z_1} \cdot \frac{\partial L}{\partial p_1} = \frac{\partial p_1}{\partial z_1} \cdot \frac{\partial z_2}{\partial p_1} \cdot \frac{\partial L}{\partial z_2}$$

$$= \overset{N \times C}{\frac{\partial L}{\partial z_2}} \cdot \overbrace{\frac{\partial z_2}{\partial p_1}}^{C \times H} \overset{T}{W_2} \cdot \underbrace{(1 \text{ if } \overset{N \times H}{z_1}[i,j] > 0 \text{ , else } 0)}_{\text{elemt-wise}} \in \mathbb{R}^{N \times H} \quad (\text{average over } N)$$

$$\frac{\partial L}{\partial W_1} = \frac{\partial z_1}{\partial W_1} \cdot \frac{\partial L}{\partial z_1} = \overset{D \times N}{X^T} \cdot \left(\overset{N \times H}{\frac{\partial L}{\partial b_1}}\right)^T \in \mathbb{R}^{D \times H}$$