CS479/679: Neural Networks
Assignment 4
Sonal Bihani
Student ID: 21032559

Question 1: Gradient of a Convolution

[15 marks] The network on the right shows a convolutional layer in a larger network. The input is $\mathbf{x} \in \mathbb{R}^{50}$. Consider the kernel (or "filter") $\mathbf{w}$, with 9 elements and a bias, so that the input current for the next layer is

$$z_j \;=\; b \;+\; (\mathbf{w} \circledast \mathbf{x})_j \qquad \text{for} \quad j = 0, \ldots, 49 \tag{1}$$

where b is the kernel's bias, and

$$(\mathbf{w} \circledast \mathbf{x})_j \;\equiv\; \sum_{i=0}^{8} w_i \, x_{j-4+i} \, .$$

Note that $x_j = 0$ and $z_j = 0$ for $j < 0$ or $j > 49$.

Suppose you are given the gradient of the loss function with respect to these input currents,

$$\nabla_{\mathbf{z}} E \;=\; \begin{bmatrix} \dfrac{\partial E}{\partial z_0} & \cdots & \dfrac{\partial E}{\partial z_{49}} \end{bmatrix}$$

Write a formula that computes $\nabla_{\mathbf{x}} E$,

$$\nabla_{\mathbf{x}} E \;=\; \begin{bmatrix} \dfrac{\partial E}{\partial x_0} & \cdots & \dfrac{\partial E}{\partial x_{49}} \end{bmatrix} .$$

Express your answer using vectors and the $\circledast$ operator, as in (2).

Answer 1:

$$z_j \;=\; b \;+\; (\mathbf{w} \circledast \mathbf{x})_j \;=\; b \;+ \sum_{i=0}^{8} w_i \, x_{j-4+i} \qquad \text{for} \quad j = 0, \ldots, 49 \tag{2}$$

Now, computing the partial derivative of the loss function E with respect to the input currents z:

$$\frac{\partial E}{\partial x_j} = \sum_{k=0}^{49} \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial x_j}$$

Here, $\frac{\partial E}{\partial z_k}$ denotes the gradient of the loss function with respect to the input current z_k, and $\frac{\partial z_k}{\partial x_j}$ denotes the partial derivative of the input current z_k with respect to the input x_j.

To compute $\frac{\partial z_k}{\partial x_j}$, we can use the expression we derived earlier for z_j, and take the derivative with respect to x_j. This yields:

$$\frac{\partial x_j}{\partial z_k} = \begin{cases} w_{j-k+4}, & \text{if } j - k + 4 \in 0, 1, \ldots, 8 \\ 0, & \text{otherwise} \end{cases}$$

$$\frac{\partial E}{\partial x_j} = \sum_{k=0}^{49} \frac{\partial E}{\partial z_k} \cdot \frac{\partial z_k}{\partial x_j} = \sum_{k=0}^{49} \frac{\partial E}{\partial z_k} \cdot w_{j-k+4}$$

Finally, we can express this formula using vectors and the $\circledast$ operator by defining w_k as the kernel w flipped horizontally and shifted by k positions (i.e., $w_k = w_{-k}$)

$$\nabla_x E = \left(\frac{\partial E}{\partial x_0}, \frac{\partial E}{\partial x_1}, \ldots, \frac{\partial E}{\partial x_{49}}\right) \quad = \left(\frac{\partial E}{\partial z} \dot{\circledast}\, w_{-4}, \frac{\partial E}{\partial z} \dot{\circledast}\, w_{-3}, \ldots, \frac{\partial E}{\partial z} \dot{\circledast}\, w_4\right) \tag{3}$$

where $\frac{\partial E}{\partial z}$ is a vector of length 50 containing the gradients of the loss with respect to each input current, and w_k is a vector of length 9 representing the flipped and shifted kernel, defined as:

$$w_k = \begin{cases} w_{-k}, & \text{if } -4 \le k \le 4 \\ 0, & \text{otherwise} \end{cases}.$$