TD

Operations in Convolutional Neural Networks

Francesca Galassi

Question 1.

Let's consider a Convolutional Neural Network (CNN) that takes as input an image of size 6 x 6, with one channel only.

The CNN has 2 convolution layers, each of them followed by a Rectified Linear Unit (ReLU) activation layer. If not said otherwise, assume that stride = 1 and padding = 0. A last fully connected layer connects previous input neurons to a single output neuron. We do not consider any bias terms.

Specifically, the layers are as follows :

1. The first convolutional layer includes 3 filters $W_i^{(1)}$ of size 3 x 3 :

$$W_1^{(1)} = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}; \quad W_2^{(1)} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}; \quad W_3^{(1)} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

2. The second convolutional layer includes one filter $W^{(2)}$ of size 3 x 3 x 3 :

$$W^{(2)}[i,j,k] = \begin{cases} 1, & \text{if } j = k = 2 \\ 0, & \text{otherwise} \end{cases} \qquad \text{for } i,j,k \in \{1,2,3\}$$

3. The fully connected layer comprises a column vector of size 4 :

$$\mathbf{w}^{(3)} = \begin{bmatrix} 6 \\ 2 \\ 0 \\ -1 \end{bmatrix}.$$

1.1 Given the input image :

$$X = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

compute the output of the first hidden layer $H_i^{(1)} = \text{ReLU}(X * W_i^{(1)})$, with $i = 1, 2, 3$.

1.2 Given the output of the first layer, $H_{1:3}^{(1)}$, compute the output of the second hidden layer $H^{(2)} = \text{ReLU}(H_{1:3}^{(1)} * W^{(2)})$, with $H_{1:3}^{(1)}[i, j, k] = H_i^{(1)}[j, k]$ pour $i \in \{1, 2, 3\}$ and $j, k \in \{1, 2, 3, 4\}$.

1.3 Given the output of the second layer, $H^{(2)}$, compute the output of the network $y = (h^{(2)} \cdot \mathbf{w}^{(3)})$, where $h^{(2)}$ is the flattened output of the second layer.

Question 2.

Let's consider a Neural Network that solves an image classification task. The output categories are : *car*, *bicycle*, *motorbike* and *bus*. The input is a grayscale image of size 20 x 20 pixels.

Answer the following questions.

2.1 If the neural network comprises a fully connected layer only, with n units, how many parameters must be optimized by the gradient descent algorithm ?
Compute the number of parameters without and with bias terms.

2.2 If the NN comprises the following layers :

1. Convolutional layer, with m filters of size 5 x 5 ;
2. Fully Connected layer with n units.

How many parameters are optimized by the gradient descent ? Compute the number of parameters without bias terms.

2.3 If the architecture comprises the following layers :

1. Convolutional layer, with m filters of size 5 x 5 ;
2. MaxPooling layer, with stride 4 ;
3. Fully Connected layer with n units.

How many parameters are optimized by the gradient descent ? Compute the number of parameters without bias terms.

2.4 If the architecture comprises the following layers :

1. Convolutional layer, with m_1 filters of size 5 x 5 ;
2. MaxPooling layer, with stride 2 ;
3. Convolutional layer, with m_2 filters of size 3 x 3 x m_1 ;
4. MaxPooling layer, with stride 2 ;
5. Fully Connected layer with n_1 units ;
6. Fully Connected layer with n_2 units ;

Consider the bias for the output layer only.
How many parameters are optimized by the gradient descent ?

Question 3. Repeat the exercises in **Question 2** for an input RGB image.