1. Explain convolutional neural network, and how does it work?

2. How does refactoring parts of your neural network definition favor you?

3. What does it mean to flatten? Is it necessary to include it in the MNIST CNN? What is the reason for this?

4. What exactly does NCHW stand for?

5. Why are there 7\*7\*(1168-16) multiplications in the MNIST CNN's third layer?

6.Explain definition of receptive field?

7. What is the scale of an activation's receptive field after two stride-2 convolutions? What is the reason for this?

8. What is the tensor representation of a color image?

9. How does a color input interact with a convolution?

Answer:

1. A convolutional neural network (CNN) is a type of neural network that is designed to work with data that has a grid-like structure, such as images or time series data. It consists of one or more convolutional layers that apply a set of learnable filters to the input data, followed by pooling layers that reduce the spatial dimensions of the output, and then fully connected layers that perform classification or regression tasks. The filters in the convolutional layers learn to identify local patterns and features in the input data, while the pooling layers help to reduce the number of parameters and improve the network's ability to generalize to new data.
2. Refactoring parts of your neural network definition can be beneficial in several ways. It can make your network easier to understand and debug, as well as easier to modify or extend for different tasks or datasets. Refactoring can also help to identify areas of the network that are redundant or unnecessary, and remove them to reduce the computational cost or improve the network's performance.
3. Flattening is the process of converting a multi-dimensional array into a one-dimensional vector. It is often used as a preprocessing step in neural networks to convert the output of a convolutional layer into a format that can be input to a fully connected layer. In the case of the MNIST CNN, flattening is necessary because the output of the convolutional layers is a 3D tensor, while the input to the fully connected layer must be a 1D vector.
4. NCHW stands for "number of samples, number of channels, height, width". It is a common format used to represent multi-dimensional arrays (tensors) in deep learning frameworks like PyTorch and TensorFlow.
5. The 7x7x(1168-16) multiplications in the MNIST CNN's third layer come from the fact that this layer uses 1168 filters, each of size 7x7, to convolve with the output of the previous layer, which has 16 channels.
6. The receptive field of a neuron in a neural network refers to the area of the input data that affects the neuron's output. It is determined by the size and stride of the filters in the convolutional layers, as well as the pooling layers that reduce the spatial dimensions of the output.
7. After two stride-2 convolutions, the receptive field of an activation increases by a factor of four. This is because each convolution with stride 2 reduces the spatial dimensions of the input by a factor of two, but increases the receptive field by a factor of two.
8. A color image is represented as a 3D tensor with dimensions (height, width, channels). The first two dimensions represent the spatial dimensions of the image, while the third dimension represents the color channels (usually red, green, and blue).
9. In a convolutional neural network, a color input is treated as a multi-channel input, with each color channel being processed separately by the network. The filters in the convolutional layers learn to identify patterns and features in each channel, and the output of the convolutional layers is a multi-channel tensor that can be passed to the fully connected layers for classification or regression tasks.