

AI6126: Homework 1

Deadline: 31 August 2020 11:59PM

Question 1: A network with the type of each layer and the corresponding output shape is given as follows

Layer (type)	Output Shape
Conv2d-1	[-1, 6, 28, 28]
ReLU-2	[-1, 6, 28, 28]
MaxPool2d-3	[-1, 6, 14, 14]
Conv2d-4	[-1, 16, 10, 10]
ReLU-5	[-1, 16, 10, 10]
MaxPool2d-6	[-1, 16, 5, 5]
Conv2d-7	[-1, 120, 1, 1]
ReLU-8	[-1, 120, 1, 1]
Linear-9	[-1, 84]
ReLU-10	[-1, 84]
Linear-11	[-1, 10]
LogSoftmax-12	[-1, 10]

The input has a shape of 1x32x32. The output shape of each layer is provided as [<ignore>, output channels, height, width]. For instance, at layer ‘Conv2d-1’, the output shape is [6, 28, 28], i.e., six feature maps of spatial size 28x28. Each conv filter and neuron of linear layer has a bias term and stride = 1.

Calculate the number of parameters for each layer and finally the total number of parameters of this network.

Question 2: Define a model in PyTorch with the architecture as given in Question 1. Start with the following constructor

```
class HelloCNN(nn.Module):
    def __init__(self):
        super(HelloCNN, self).__init__()

    def forward(self, x):
```

Question 3: Please answer the following questions:

- i) Explain the difference between regression and classification.
- ii) You need to train a neural network that predicts the age of a person. Is this a regression or classification problem?
- iii) Why do we need a validation set?

Question 4: Let us consider the convolution of single-channel tensors $\mathbf{x} \in \mathbb{R}^{4 \times 4}$ and $\mathbf{w} \in \mathbb{R}^{3 \times 3}$

$$\mathbf{w} \star \mathbf{x} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 10 & 10 & 0 & 0 \\ 10 & 10 & 0 & 0 \\ 10 & 10 & 0 & 0 \\ 10 & 10 & 0 & 0 \end{pmatrix}$$

Perform convolution as matrix multiplication by converting the kernel into sparse Toeplitz circulant matrix. Show your steps.

Question 5: Why might we prefer to minimize the sum of absolute residuals (L1 loss) instead of the residual sum of squares for some data sets (L2 loss)? (*Hint:* What is one of the flaws of least-squares regression?)