

1. Consider a simple linear classifier with  $m$  classes. The input to the classifier is vector  $x \in \mathbb{R}^n$ . The first layer is linear layer whose output is  $Wx$ , and  $W \in \mathbb{R}^{m \times n}$  is a parameter matrix. Then the cross-entropy loss function denoted by  $L$  is applied to the output (that is, first applying softmax to the output and applying negative log-likelihood). Suppose that the current input is  $x$ , and the ground truth label is given by  $y \in \{1, \dots, m\}$ . Find the expression for

$$\frac{dL}{dW}$$

Note that the answer should have the same shape as  $W$ .

2. Consider a network with a convolutional layer followed by ReLU activation. The input map is denoted by  $x$  given by

$$x = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 0 \\ 0 & -1 & -2 \end{bmatrix}$$

The parameter for convolutional filter is denoted by  $W$  which has spatial dimension of height 2 and width 2 (no bias), with no padding and stride of 1. The current value of  $W$  is given by

$$W = \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}$$

Let the output of ReLU activation be  $y$ . Suppose that the upstream gradient of loss  $L$  with respect to  $y$  is given by

$$\frac{dL}{dy} = \begin{bmatrix} -3 & 2 \\ 0 & 1 \end{bmatrix}$$

Find the derivative of  $L$  with respect to parameter  $W$ , that is

$$\frac{dL}{dW}$$

Note that the answer should have the same shape as  $W$ .

3. Suppose input tensor of size  $(C, H, W)$  is fed into a convolutional network, where  $C$ ,  $H$ ,  $W$  represent channel, height and width. Suppose the forward computation goes through the following layers in sequence:
- (a) conv layer, number of filters: 8, spatial filter size:  $7 \times 7$ , padding 2, stride 2
  - (b) max pool, filter size  $2 \times 2$ , padding 0, stride 2
  - (c) conv layer, filters: 16, spatial filter size:  $5 \times 5$ , padding 1, stride 2
  - (d) conv layer, filters: 16, spatial filter size:  $3 \times 3$ , padding 1, stride 1
  - (e) max pool, filter size  $2 \times 2$ , padding 0, stride 2

Suppose the input shape is given by  $(4, 37, 37)$ . What is the shape of output tensor? Write in  $(C, H, W)$  format.

4. Consider the Inception module as shown in the figure. The input is replicated and fed along four paths, where the input channel size is 192. The modules in each path has the following configuration

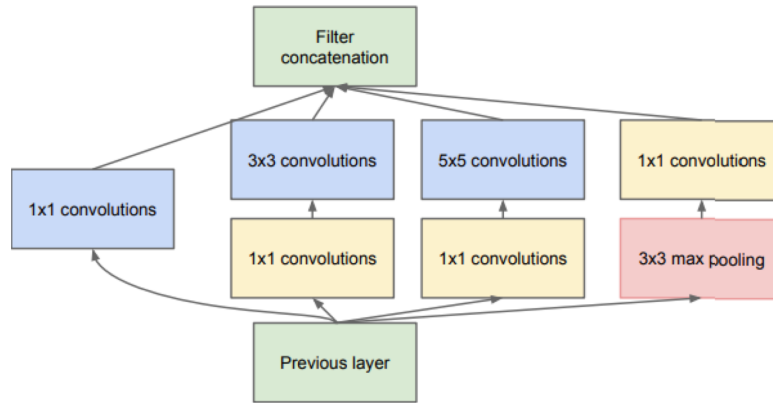


Figure 1: Inception module

- (a) 1x1 conv (spatial dimension of filter: 1x1, number of filters: 64)
- (b) 1x1 conv (spatial dimension of filter: 1x1, number of filters: 96) → 3x3 conv (spatial dimension of filter: 3x3, number of filters: 128)
- (c) 1x1 conv (spatial dimension of filter: 1x1, number of filters: 16) → 5x5 conv (spatial dimension of filter: 5x5, number of filters: 32)
- (d) max pool → 1x1 conv (spatial dimension of filter: 1x1, number of filters: 32)

Find the number of parameters in the Inception module. Assume that there is no bias parameter in all of the convolutional layers.