

Advanced Topics in Machine Learning – Assignment # 5 Universität Bern

Meiguang Jin (jin@inf.unibe.ch)
Mehdi Noroozi(noroozi@inf.unibe.ch)

Due date: 24/05/2017

In this assignment you need to write down the answers to the question. We prefer pdf format (over docx) because it is universally supported on all platforms. Please also indicate your name on top of the first page in your pdf submission.

Maximum grade: 4 Points

1. AlexNet is the most famous convolutional neural network (CNN) in the computer vision community. The architecture of the network is shown in Table 1. The table only shows convolutional, pooling, and fully connected layers. There is a ReLU layer after all the convolutional and fully connected layers except for `fc8`. There is also a normalization layer after `pool1` and `pool2` layer. A Dropout layer is also used after `fc6`, `fc7` to avoid overfitting. These layers are not shown in the table for the sake of simplicity. Suppose we train AlexNet with $W \times H$ pixel images, where $W = H = 200 + \text{two right most digits of your student ID (8 digits)}$. For example if your student ID is 14-141-512, then $W = H = 212$.

Answer the following questions

- (a) **[.5 point]** What are the spatial dimensions of the outputs of all the layers from `conv1` to `pool5`?
- (b) **[.5 point]** How many parameters does the network have? Show how you calculate the number of parameters of each layer separately. Which layer consists of most of the parameters?

Name	conv1	pool1	conv2	pool2	conv3	conv4	conv5	pool5	fc6	fc7	fc8
Type	conv	pool	conv	pool	conv	conv	conv	pool	fc	fc	fc
OutCh	96	96	256	256	384	384	256	256	4096	4096	1000
Kernel	11	3	5	3	3	3	3	3	-	-	-
Stride	4	2	2	2	1	1	1	1	-	-	-
Pad	0	0	2	0	1	1	1	0	-	-	-

Table 1: The AlexNet architecture

- (c) **[.5 point]** Calculate the receptive field of the neurons of the pool5 layer.
- (d) **[.5 point]** Suppose that you are given a pretrained AlexNet network. We feed the network with a new image and take the output of intermediate layers as features. Compute what input image sizes can be fed to the network, if we want to take the fc6 layer features. Is there such a limitation if we want to use convolutional layer activations as features?

2. Answer the following questions

- (a) **[0.5 point]** Given an input I with size $1 \times C \times H \times W$ and kernel K with size $C \times N \times M$, write the explicit formula for a convolution Z_{ij} with stride s .
- (b) **[1.5 points]** In Recurrent Neural Networks training, one potential issue is *gradient vanishing*, especially with long-term dependencies. Explain why this happens. Justify your answer mathematically.