

National University of Singapore
School of Computing
CS3244: Machine Learning
Tutorial 8

Deep Learning

Colab Notebook : Deep Learning

Typo in Slides 10.a - Figure 1 shows the correction.

Correction

What are the Kernel Size, Stride, Padding?

$\{height \times width\}$ $\dim \mathbf{x} = \{2 \times 6\}$

$\mathbf{W} = \begin{pmatrix} -1 & 0 \\ -1 & 1 \\ 0 & 1 \end{pmatrix}$ $\mathbf{x} = \begin{pmatrix} 9 & 9 & 9 & 3 & 3 & 4 \\ 9 & 9 & 3 & 5 & 5 & 8 \end{pmatrix}$

$h_p/2$ (padding) w_s (stride)

$\dim \mathbf{y} = \{4 \times 3\}$

$\mathbf{y} = \mathbf{W} * \mathbf{x} = \begin{pmatrix} 0+0+9 & 0+0+3 & 0+0+4 \\ 0+0+9 & 0-6+5 & 0+1+8 \\ -9+0+0 & -9+2+0 & -3+3+0 \\ -9+0+0 & -3+0+0 & -5+0+0 \end{pmatrix}$

Hyperparameters

- Kernel size $\kappa = \{3 \times 2\}$
- Padding $\mathbf{p} = \{(2+2) \times 0\}$
- Stride $\mathbf{s} = \{1 \times 2\}$

Chosen manually, or automatically with [hyperparameter tuning](#)

$$\dim \mathbf{y} = \left\{ \left(\frac{h_x + h_p - h_\kappa}{h_s} + 1 \right) \times \left(\frac{w_x + w_p - w_\kappa}{w_s} + 1 \right) \right\}$$

$$= \left\{ \left(\frac{2+4-3}{1} + 1 \right) \times \left(\frac{6+0-2}{2} + 1 \right) \right\}$$

NUS CS3244: Machine Learning 4

Figure 1: Convolution Operation and Output

1. **CNN** You are given an image $\mathbf{x} \in \mathbb{R}^{4 \times 4}$ and a kernel $\mathbf{W} \in \mathbb{R}^{3 \times 3}$. **No** extra padding is added.
 - (a) Get the output feature map from this convolution and write down its output dimensions if the kernel moved with a stride of 1×1 and we used ReLU. No Pooling operation required.
 - (b) Now, let's say you have access to a very deep, large CNN model. We feed a single image to the network. Each image has $3(C)$ channels (RGB) with a height and width of 224×224 ($H = 224, W = 224$). Here our input is a 3-dimensional tensor ($H \times W \times C$). The first layer of the big CNN is a Convolutional Layer with 96 (C_1) kernels of has a height of 11 and a width of 11 and each kernel has the same number of channels as the input image. Stride is 4×4 and no padding is used. Calculate the output size after the first Convolutional Layer.

0.5	0.2	0.1	0.7
0.1	0.6	0.9	0.5
0.0	0.8	0.2	0.7
0.2	0.4	0.0	0.4

0.1	0.2	0.6
0.4	0.3	0.5
0.9	0.8	0.7

Figure 2: (left) The image \mathbf{x} . (right) The kernel \mathbf{W}

- (c) In most of Deep Learning libraries such as PyTorch, and TensorFlow, Images are fed together as a batch(B). B can take values such as 8, 16, 32, 64. Comment on output shape if we fed the large CNN with a batch of images in part (b). What are the advantages of using a batch of images rather than a single image?
- (d) In a CNN, why is it good design choice to stack **TWO** Convolutional layers, each with a 3×3 kernel, one after the other on an image instead of a **SINGLE** Convolutional Layer with a 5×5 kernel? The Convolutional layers have 1×1 stride. (**Hint:** Think about parameter count and power of representations)
2. **Quick Discussions.** Describe what deep learning models/techniques will you use to tackle the following problems.
- You need a model to determine the sentiment of sentences.
 - You need a model to do translation between two languages.

You can refer to the following reading materials to learn more:

1. Convolutional Neural Networks for Sentence Classification
2. Residual Networks Behave Like Ensembles of Relatively Shallow Networks
3. Attention Is All You Need