1. The first hidden layer of a convolution neural network (CNN) has a convolution layer consisting of two feature maps with filters $\boldsymbol{w_1}$and $\boldsymbol{w_2}$ and biases = 0.1, and neurons having sigmoid activation functions, and a pooling layer with a pooling windows of size 2x2:

$$\boldsymbol{w}_1 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \text{ and } \boldsymbol{w}_2 = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

The input layer is of 6x6 size and receives an input image $\boldsymbol{I}$:

$$\boldsymbol{I} = \begin{pmatrix} 0.7 & 0.1 & 0.2 & 0.3 & 0.3 & 05 \\ 0.8 & 0.1 & 0.3 & 0.5 & 0.1 & 0.0 \\ 1.0 & 0.2 & 0.0 & 0.3 & 0.2 & 0.7 \\ 0.8 & 0.1 & 0.5 & 0.6 & 0.3 & 0.4 \\ 0.1 & 0.0 & 0.9 & 0.3 & 0.3 & 0.2 \\ 1.0 & 0.1 & 0.4 & 0.5 & 0.2 & 0.8 \end{pmatrix}$$

   a. Find the outputs at the first convolution layer if
      i. padding = 0 and strides = $(1, 1)$.
      ii. padding = 1 and strides = $(2, 2)$.
   b. Find the outputs at the first pooling layer for Part (a), assuming strides of $(2, 2)$ and pooling is
      i. max pooling
      ii. mean pooling

2. Given '3wolfmoon.jpg' color image of size 639×516.
   a. Initialize weights and biases of a convolutional layer with two kernels of size 9×9. Note that the input image is in color and has three channels.
   b. Display the feature maps at the convolution layer, assuming sigmoid activation functions. Use VALID padding and strides = 1.
   c. Display the outputs of a mean pooling layer with a pooling window size 5×5 and strides = 5.

3. Design a CNN with one hidden layer to recognize digit images in MNIST database:

   http://yann.lecun.com/exdb/mnist/

   The convolution layer consists of 25 filters of dimensions 9x9 and the pooling layer has a pooling window size 4x4. Assume VALID padding and default strides for both convolution

and pooling layer. Train the network with mini batch gradient decent learning with learning factor $\alpha = 10^{-3}$ and batch size = 128.

Plot

a. The training and test errors against learning epochs.
b. Final filter weights
c. Feature maps at the convolution and pooling layers for a representative test pattern
d. Repeat training by introducing decay parameter $\beta = 10^{-6}$ and momentum term with $\gamma = 0.5$, and compare the learning curves