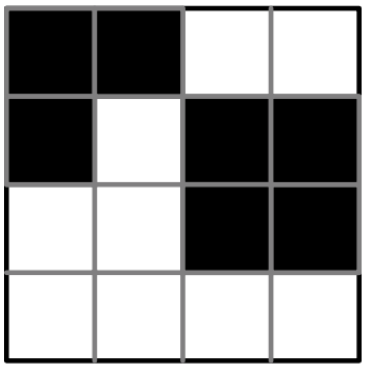
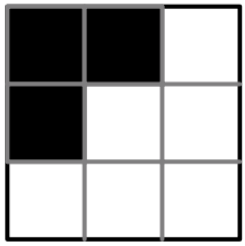
Name: _____

Convolutional News Network

3. (10 points) (a) Consider the following image (on the left) and filter (on the right):





Consider what results from filtering this image with this filter, assuming that the input image is padded with zeros, and using a stride of 1. To compute the output value of a particular pixel (i, j), apply the filter with its center on pixel (i, j) of the input image.

Assume dark pixels have a value of 1 and light pixels have a value of -1.

i. What is the output value for the top-left image pixel (that is, the pixel with indices (1, 1) in one-based indexing)?

-2

ii. What element of the output image will have the highest value? (Assume the rows and columns of the image are numbered starting with 1.)

3, 1

(b) If we used 5 different filters with size 3×3 and stride 1 on this image, what would the dimensions of the resulting output be?

Solution: $4 \times 4 \times 5$

Using the formula for output size from the course CNN notes (page 70):
output size (in 1 dimension) = ceil((input size + 2*padding - (kernel size - 1))/stride)
= 4 + 2*1 - 2 = 4
So we get a 4x4 output, and there are 5 outputs since there are 5 filters —> 4 x 4 x 5.
Note that it is not made explicitly clear that the padding is of size 1 in the question. It can be inferred/assumed because for a 3x3 kernel, padding of size 1 is needed in order for the center of the kernel to fall on the center of every pixel of the (original) input.

(c) What would be the result of applying max-pooling with size $k = 2$ and stride 2 on the original, unfiltered image above?

i. What are the dimensions of the resulting image?

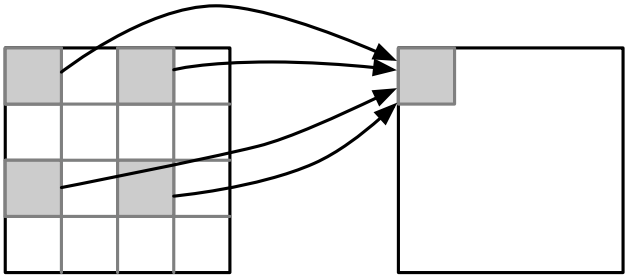
2×2

ii. Draw the actual image with numerical values for each pixel in the space below.

Solution:

```
 1 1
-1 1
```

(d) Dana has an idea for a new kind of network called a ModConv NN. If the network is $n \times n$, we will use a filter of size n/k (assume k evenly divides n). To compute entry (a, b) of the resulting image, we apply this filter to the "subimage" of pixels (i, j) from the original image, where $i \bmod k = a$ and $j \bmod k = b$.



i. Could we train the weights of a ModConvNN using gradient descent? Explain why or why not.

Solution: Sure. Just another parametric model.

ii. What underlying assumption about patterns in images is built into a regular convolutional network, but not this one?

Solution: This one does not encode the fact that nearby groups of pixels work together to encode information (that there is spatial locality of useful patterns in an image).