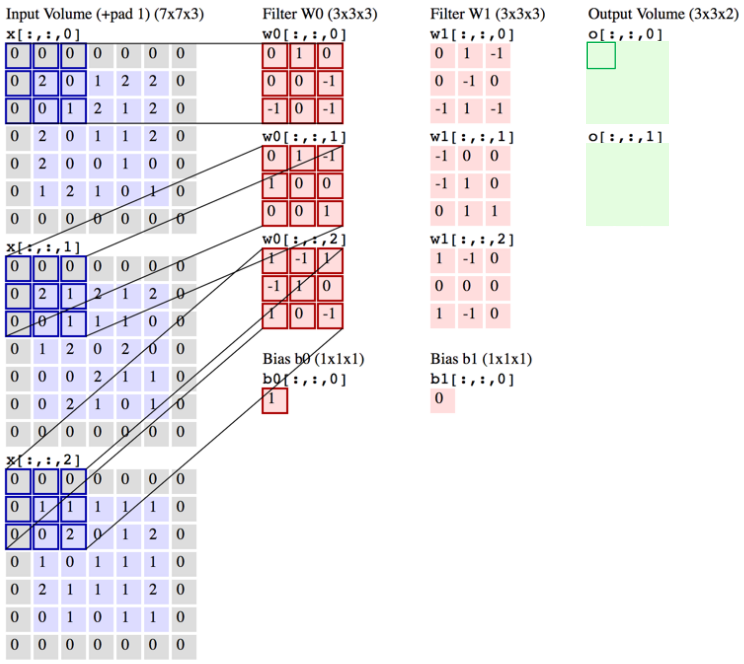
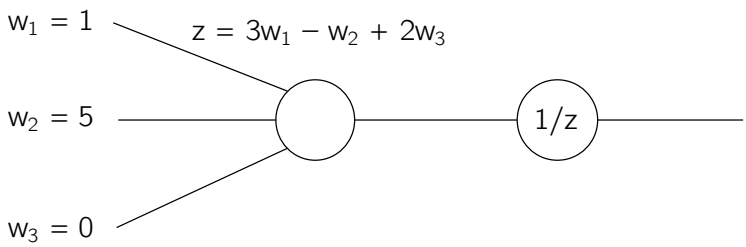
Midterm 2 Practice Problems *(find and work with a partner)*

1. *Using CNN filters.* In the example below, use the 0th filter to fill in the value of the green square in the output (and optionally compute other values). Why is the depth of the output 2 instead of 3?



2. *Backpropagation.* Consider the following network diagram representing the function

$$f(w_1, w_2, w_3) = \frac{1}{3w_1 - w_2 + 2w_3}$$

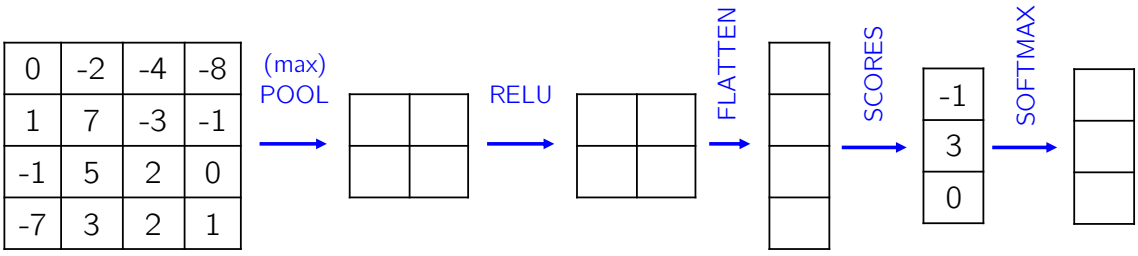


   (a) Given the initial values for w_1, w_2, w_3, compute the forward pass to find the function value.

   (b) Assuming a "1" on the gradient coming back, compute the backward pass to find the gradient of the output with respect to each input.

   (c) If the learning rate $\alpha = 1$, how would w_1, w_2, w_3 be updated during a step of gradient descent?

3. *High-level CNN.* Say I have the following output of a CONV layer on the left. Assume no bias terms throughout.



(a) If my original input was also 4×4 and I used one convolutional filter with size 3×3 (no bias), how much zero padding would I need? How many parameters would I need to learn just for this CONV layer?

(b) Fill in the steps POOL (2×2 with stride 2), RELU, FLATTEN.

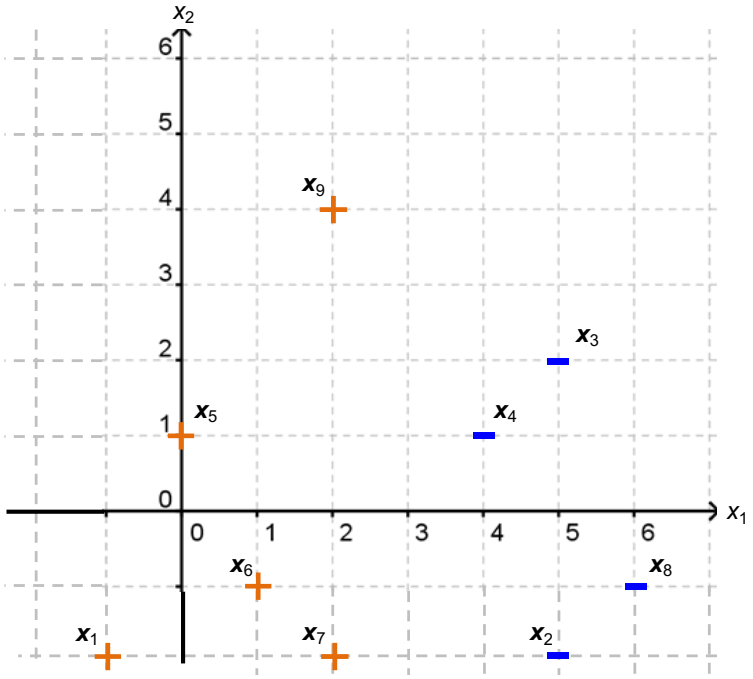
(c) If there is a single dense (fully connected) layer between the flattened output and the scores, how many parameters do we need to learn just for this layer?

(d) Say the scores are as given above for three potential labels $\hat{y} \in \{1, 2, 3\}$. Compute the SOFTMAX function to obtain a probability distribution over these three classes. What would you choose for the predicted label $\hat{y}$?

(e) If the true class was in fact $y = 2$, what is the cross-entropy loss?

(f) In the input had been a matrix of zeros, what would the scores be? What would the probability distribution (output of SOFTMAX) be?

4. *SVM and Perceptron.* Let $K = 4$ (initial dataset size). We will iteratively add points in order of their indices (not randomly). Run the incremental SVM optimization algorithm – at each stage, write out S (the set of points being considered), the support vectors, and which α values end up being 0. At the end, what is the equation of the separating hyperplane?



Round 1:

- $S =$
- Support vectors:
- α's that are 0:

Round 2:

- $S =$
- Support vectors:
- α's that are 0:

Round 3:

- $S =$
- Support vectors:
- α's that are 0:

Round 4:

- $S =$
- Support vectors:
- α's that are 0:

5. For the dataset above, run the perceptron algorithm using the points in order. What is the equation of the separating hyperplane in this case?