CSC321H5 Homework 5.

Deadline: Thursday, March. 26, by 9pm

Submission: You must submit your solutions as a PDF file through MarkUs. You can produce the file however you like (e.g. LaTeX, Microsoft Word, scanner), as long as it is readable.

Late Submission: Please see the syllabus for the late submission criteria.

Question 1. – 8 pts

Part (a) – 6 points

Explain the difference between the “stride”, “padding” and “output_padding” settings in a Transpose Convolution operation.

Part (b) – 2 points

Explain why autoencoders tend to produce blurry images.

Question 2. – 4 pts

The website https://isthisacat.com/ allows you to upload an image, and provides a prediction of whether or not the image is a cat. This website uses a neural network to make its predictions.

Suppose we wish to mount an adversarial attack: we have a picture of a cat, and want to fool the service into thinking that it is not a cat.

Part (a) – 2 points

You don’t have access to the neural network weights and biases that the website uses, so we will need to mount a black-box attack. How will you train your neural network? In particular, what image data will you use (0pt) and how will you generate the target labels (1pt)? What loss will you use (1pt)?

Part (b) – 2 points

Now, let’s suppose that we start with an image of a cat. How will you use your model to tune the image, so that the website will not recognize it to be a cat? In particular, what values (i.e. parameters) will you optimize (1pt)? What loss function could you use (1pt)?

Question 3. – 10 pts

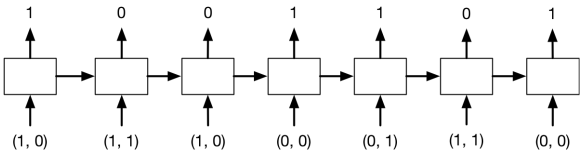
In this problem, you will implement a recurrent neural network which implements binary addition. The inputs are given as binary sequences, starting with the *least* significant binary digit. (It is easier to start from the least significant bit, just like how you did addition in grade school.) The sequences will be padded with at least one zero on the end. For instance, the problem

[100111 + 110010 = 1011001]

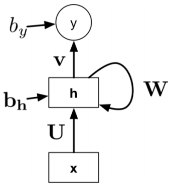
would be represented as:

- **Input 1:** 1, 1, 1, 0, 0, 1, 0
- **Input 2:** 0, 1, 0, 0, 1, 1, 0
- **Correct output:** 1, 0, 0, 1, 1, 0, 1

There are two input units corresponding to the two inputs, and one output unit. Therefore, the pattern of inputs and outputs for this example would be:



Design the weights and biases for an RNN which has two input units, three hidden units, and one output unit, which implements binary addition. All of the units use the hard threshold activation function ($f(x) = 1$ if $x > 0$ and 0 otherwise). In particular, specify weight matrices $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$, bias vector $\mathbf{b_h}$, and scalar bias b_y for the following architecture:



Hint: In the grade school algorithm, you add up the values in each column, including the carry. Have one of your hidden units activate if the sum is at least 1, the second one if it is at least 2, and the third one if it is 3.

Question 4. – 8 pts

Part (a) – 6 points

Suppose we want to use a recurrent neural network to predict the sentiment expressed in a piece of text.

List and explain two potential advantages and two potential disadvantages of using a *character-level* recurrent neural network, as opposed to a *word-level* recurrent neural network.

Part (b) – 2 points

We saw in lecture that GloVe embeddings have human biases encoded in their structure. Why are GloVe embeddings biased? In other words, where do those biases come from?