ECE/CS 559 - Spring 2020 - Final Exam

Full Name: **ID Number:**

Justify all answers. Answers without justification will not receive credit.

No neural network libraries as usual.

Q0. (5 points) Attach the email indicating you have completed the instructor and TA evaluations.

Q1. Consider the RNN given by the state update equation

$$s_t = \tanh(wx_t + us_{t-1}), \tag{1}$$

where $s_t, w, x_t, u \in \mathbb{R}$ are all one-dimensional scalars. We will use this RNN for time-series estimation. Specifically, we will be given most recent k values of a time series and we will predict the next value.

Set the initial seed state $s_0 = 0$. Unroll the RNN $k > 1$ times to get the input output relationships:

$$s_t = \tanh(wx_t + us_{t-1}),\ t = 1, \ldots, k. \tag{2}$$

In this scenario, we are given the most recent values $x_1, \ldots, x_k$ of the time series, and we are trying to predict x_{k+1}. We will use s_k as our prediction for x_{k+1}. The prediction error will be measured in the mean squared sense

$$E = \frac{1}{2}(s_k - x_{k+1})^2 \tag{3}$$

(a) (10 points) Write down $\frac{\partial E}{\partial u}$ and $\frac{\partial E}{\partial w}$ for the special case of unrolling the RNN $k = 2$ times.

(b) (15 points) Write down a formula for $\frac{\partial E}{\partial u}$ and $\frac{\partial E}{\partial w}$ for a general k.

Q2. (30 points) Consider the time series $z_t = \sin(\cos(t/10) + 2\cos(t/5))$, $t = 1, 2, \ldots$. Pick some $k \in \{5, \ldots, 10\}$. Begin with some random u and w. Train the RNN in Q1 with the training set

$$\begin{aligned}
[x_1, \ldots, x_k] &= [z_1, \ldots, z_k],\ x_{k+1} = z_{k+1} \\
[x_1, \ldots, x_k] &= [z_2, \ldots, z_{k+1}],\ x_{k+1} = z_{k+2} \\
[x_1, \ldots, x_k] &= [z_3, \ldots, z_{k+2}],\ x_{k+1} = z_{k+3} \\
&\vdots \\
[x_1, \ldots, x_k] &= [z_{200}, \ldots, z_{k+199}],\ x_{k+1} = z_{k+200}
\end{aligned}$$

Plot $z_{k+1}, \ldots, z_{k+200}$ (on the y-axis) with $k+1, \ldots, k+200$ as the x-values. These are the actual values of the time series. On the same graph, plot the predictions of your RNN. In the beginning, the predictions should be far-off (as u and w are untrained), but they should progressively get better. In another graph, plot $k+1, \ldots, k+200$ versus the MSEs. Comment on the results.

Q3. Consider a CNN whose input is a 20×20 image. The first layer is a convolution layer, consisting of 8 filters of size 3×3 each. The second layer is a max pooling layer with window size 2×2 and a stride of 2. The last layer is a fully-connected layer with 10 neurons.

(a) (5 points) How many free parameters are there to optimize in this network? Justify your answer. Suppose there are no biases.

(b) (5 points) Suppose that the linear activation function is used in all layers. Does the network provide a linear function of its inputs?

(c) (5 points) Repeat part (b) for a network without the first convolution layer (max pooling layer as the first layer and fully connected layer as the second layer).

(d) (5 points) Repeat part (b) for a network without the max-pooling layer (convolution layer as the first layer and the fully connected layer as the second layer).

Q4. Consider a GAN with the generator

$$G(z) = \tanh\left(\sum_{i=1}^{2} w_i^G \tanh(u_i^G z)\right)$$

and the discriminator

$$D(x) = \frac{1}{2}\left(1 + \tanh\left(\sum_{i=1}^{2} w_i^D \tanh(u_i^D x)\right)\right)$$

(a) (10 points) Write down the explicit formula to update u_1^D during the discriminator update.

(b) (10 points) Write down the explicit formula to update w_2^G during the generator update.

(c) (For the adventurous) Train the GAN (possibly with more neurons in the hidden layers' of the discriminator and the generator, including biases) for some non-uniform input distribution to your liking and some noise distribution. Demonstrate that the generator distribution matches the input distribution at convergence.