

Deep Learning CS583 Fall 2021

Quiz 3

December 1, 2021

Instructor: Jia Xu

Student name: _____

Student ID: _____

Student email address: _____

- **Read these instructions carefully**
- Fill-in your personal info, as indicated above.
- You have 24 hours.
- There are five items in one question. Each item worths the same (1 point).
- Both computer-typed and hand-writing in the very clear form are accepted.
- This is an open-book test.
- You should work on the exam only by yourself.
- Submit your PDF/Doc/Pages **by 12:30 Dec 3rd** on Canvas under Final exam.

good luck!

1 Question

- What is the advantage and disadvantage of attentional models compared to RNNs.

Choose one correct answer from four candidates:

- In practice, what is the most accurate description for activation functions (such as Sigmoid, Sum, Tanh, ReLU) used in neural networks?
 1. They must be differentiable.
 2. They can be non-differentiable, but only for a small number of points.
 3. They can be any continuous functions.
 4. They must be non-linear to be learnable.
- Given a neural network with N input nodes, no hidden layers, one output node, with entropy loss and sigmoid activation functions, which of the following algorithms (with the proper hyper-parameters and initialization) can be used to find the global optimum?
 1. Stochastic Gradient Descent
 2. Batch Gradient Descent
 3. Mini-Batch Gradient Descent
 4. All of the above
- You want to train a neural network to predict the next 30 daily prices using the previous 30 daily prices as inputs. Which model selection and explanation make the most sense?
 1. A fully connected deep feed-forward network because it considers all input prices in the hidden layers to make the best decision.
 2. A single one-directional RNN because it considers the order of the prices, and the output length is the same as the input length.

3. A bidirectional RNN because the prediction benefits from future labels.
 4. A one-directional encoder-decoder architecture can generate a sequence of future prices based on all historical input prices.
-
- Draw the computational graph of a one-hidden layer feed-forward neural network and write the derivatives of each variable in the backpropagation.