Return to Assessments page

Name: _____
Score: _____ / _____

Worksheet 08 - RNNs and regularization

Part 1: Regularization

Gradient decay

Given a deep neural network with depth d, one hidden neuron each, and sigmoid $\sigma(\cdot)$ activation functions representing a function $f : \mathbb{R} \to \mathbb{R}$. The loss function is given by $\ell(x, y) = (x - y)^2$. How does the gradient of the loss with respect to the first weight w_1 scale in the number of layers and the maximum possible gradient of the activation $\nu := \max_x \frac{d\sigma}{dx}(x)$ for fixed x?

- A. $\mathcal{O}(\nu d)$
- B. $\mathcal{O}(\nu^d)$
- C. $\mathcal{O}(\nu \log d)$
- D. $\mathcal{O}(\sqrt[d]{\nu})$

Answer Point Value: 1.0 points
Answer Key: B

Dropout

When training a NN with a dropout layer with rate p and unscaled weights during the training, what can be said about the weights during test time?

- A. **They are used just as learned in the training process.**
- B. **They are rescaled with p.**
- C. **They are randomly set to 0 with probability p.**
- D. **They are randomly set to 0 with probability $(1 - p)$.**

Answer Point Value: 1.0 points
Answer Key: B

Regularization

Which of the following regularizations encourage sparsity in the solution?

- A. **Dropout**
- B. **L0 regularization**
- C. **L2 regularization**
- D. **None of dropout, L0, L2**
- E. **All of dropout, L0, L2**

Answer Point Value: 1.0 points

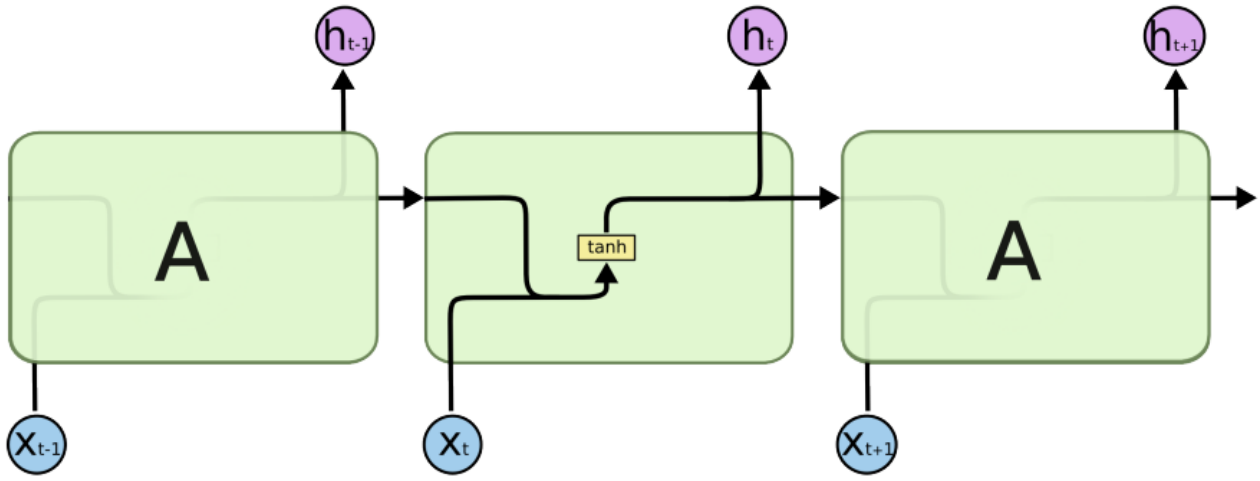
Answer Key: B

Part 2: Recurrent networks

RNNs

In the general case, which of the RNN models need the least computational time for a training step?

A simple RNN is sketched below.



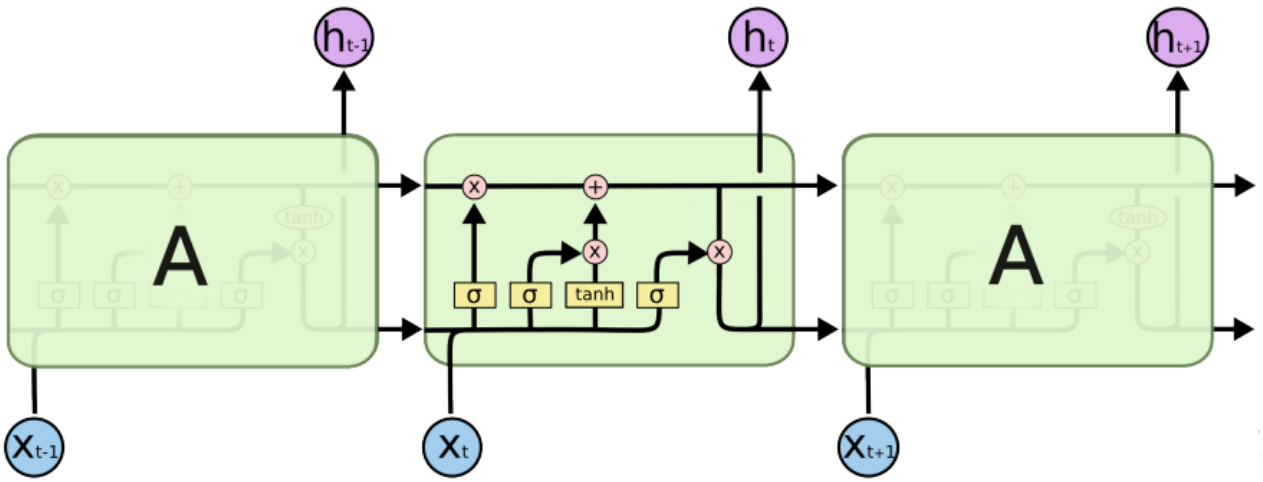
- A. **Simple RNN (as in image)**
- B. **Simple RNN with teacher forcing**
- C. **LSTM**
- D. **The computational time is comparable across the models**

Answer Point Value: 1.0 points

Answer Key: B

LSTM

We want to initialize a modified LSTM layer as shown in the image so that at the beginning of the training it will behave in the same way as an untrained simple RNN layer (see image of previous question) up to arbitrary accuracy. How do we need to initialize the parameters of the different parts of the layer? Refer to the slides for the naming of the different weight and biases.



Modified LSTM

Forget gate weights W_f:

- A. **Small random values**
- B. **Small positive values**
- C. **Small negative values**
- D. **Large positive values**
- E. **Large negative values**
- F. **Zeros**

Answer Point Value: 1.0 points
Answer Key: F

Forget gate biases b_f:

- A. **Small random values**
- B. **Small positive values**
- C. **Small negative values**
- D. **Large positive values**
- E. **Large negative values**
- F. **Zeros**

Answer Point Value: 1.0 points
Answer Key: E

Input gate weights W_i

- A. **Small random values**
- B. **Small positive values**
- C. **Small negative values**
- D. **Large positive values**
- E. **Large negative values**
- F. **Zeros**

Answer Point Value: 1.0 points
Answer Key: F

Input gate biases b_i

- A. **Small random values**
- B. **Small positive values**
- C. **Small negative values**

- D. **Large positive values**
- E. **Large negative values**
- F. **Zeros**

Answer Point Value: 1.0 points
Answer Key: D

New candidates weights W_C

- A. **Small random values**
- B. **Small positive values**
- C. **Small negative values**
- D. **Large positive values**
- E. **Large negative values**
- F. **Zeros**

Answer Point Value: 1.0 points
Answer Key: A

New candidates biases b_C

- A. **Small random values**
- B. **Small positive values**
- C. **Small negative values**
- D. **Large positive values**
- E. **Large negative values**
- F. **Zeros**

Answer Point Value: 1.0 points
Answer Key: A

Output gate weights W_o

- A. **Small random values**
- B. **Small positive values**
- C. **Small negative values**
- D. **Large positive values**
- E. **Large negative values**
- F. **Zeros**

Answer Point Value: 1.0 points
Answer Key: F

Output gate biases b_o

- A. **Small random values**
- B. **Small positive values**
- C. **Small negative values**
- D. **Large positive values**
- E. **Large negative values**
- F. **Zeros**

Answer Point Value: 1.0 points
Answer Key: D

Part 3

Please state the names of all the students you worked with on this assignment:

Answer Point Value: 0.0 points
Model Short Answer: --------