

1. While training a vanilla RNN, the ML engineer finds that the weights keep growing with each epoch. Which of the following could be the possible reasons and solutions:

- A. High learning rate. Try lowering learning rate.
- B. Not enough data. Get more training data.
- C. Bad architecture. Use LSTM
- D RIGHT. Exploding gradients. Try clipping gradient.

- 
- Accepted answer is A and D.

I think the 'learning rate' option will be right only if 'the **magnitude of** weights keep growing'. Because high learning rate causes for instability. Means 'the **magnitude of** weights keep growing' but sign will keep on changing to negative and positive.

---

2. Which of the following statements are true

- A. The forward pass of Vanilla RNNs is cheaper than that of GRUs and LSTMs
- B. Amongst GRUs, LSTMs and Vanilla RNNs, only LSTMs have a separate memory cell
- C. Usually, LSTMs can compute deeper sequences (without vanishing gradients) compared to GRUs
- D RIGHT. All of the above

3. What common property of AlexNet, LSTMs and ResNet helps in training?

- A. The number of their layers.
- B. The number of parameters
- C RIGHT. They have multiple pathways for gradient backflow which helps in backprop
- D. All of the above

LOOK LIKE THE QUESTION IS WRONG. AlexNet is odd here.

A company is trying to automate case reports for MRI scans. The scans are videos of a beating heart at a particular cross section (slice). The automatic report is supposed to give a diagnosis amongst 5 different conditions (one of which is a "normal heart"). The videos are of 30 frames each. Each frame is a 227x227 grayscale image. Answer the following

4. Which is the most appropriate classification for the type of RNN to be used for this problem?

- A. One to One
- B. One to Many
- C RIGHT. Many to One
- D. Many to Many

5. If we are to use the full, unprocessed video as the input to the RNN, which of the following are true?

- A RIGHT. The unrolled RNN will have a depth of 30 in time.

- **B RIGHT.** There are 30 sequential inputs to the RNN each of size  $227 \times 227$
- **C.** There is only 1 non-sequential input to the RNN of size  $227 \times 227 \times 30$
- **D.** The RNN can be a deep RNN with 30 CNN like units

Accepted Answers:

- The unrolled RNN will have a depth of 30 in time.
- There are 30 sequential inputs to the RNN each of size  $227 \times 227$
- The RNN can be a deep RNN with 30 CNN like units

The input is fixed here. So We can use CNN here. That means 'deep RNN' will work. Technically that option is correct. But looks like the option is odd.

The ML engineer in the company decides that using the full, unprocessed video is too expensive for an RNN. So, she decides to encode the image by using a CNN architecture similar to Alexnet. The final, fully connected layer in her architecture has size 50. She uses this embedding (final layer) as the input to the RNN for each frame. She also uses a single hidden layer with 100 neurons. Answer the following questions.

I assume the input size is  $(50, 1)$   
size of hidden layer is  $(100, 1)$

We know the equations:

- $h_t = g(W_{hh} h_{t-1} + W_{xh} x_t + b_h)$
- $y_t = g^*(W_{yh} * h_t)$

We know the following matrix sizes:

- $h_t$  and  $h_{t-1}$  is  $(100, 1)$
- $x_t$  is  $(50, 1)$

The matrix sizes in first equation:  $(100, 1) = (m_{hh}, n_{hh}) (100, 1) + (m_{xh}, n_{xh}) (50, 1) + (m_{bh}, n_{bh})$

So  $W_{hh}$  is  $(100, 100)$ ,  $W_{xh}$  is  $(100, 50)$  and  $b_h$  is  $(100, 1)$

The matrix sizes in Second equation:  $(100, 1) = (m_{yh}, n_{yh}) * (100, 1)$

So  $W_{yh}$  is (no data to identify row size , 100)

## 6. What is the size of the matrix $W_{hh}$ ?

- **A.**  $5000 \times 5000$
- **B.**  $50 \times 50$
- **C RIGHT.**  $100 \times 100$
- **D.**  $30 \times 30$

## 7. What is the size of the matrix $W_{xh}$ ?

- **A.**  $50 \times 100$
- **B RIGHT.**  $100 \times 50$
- **C.**  $100 \times 100$
- **D.**  $50 \times 50$

**8. What is the size of the bias matrix for the recurrent unit?**

- A. 50x1
- **B RIGHT.** 100x1
- C. 30x1
- D. 5000x1

**9. What is the size of the matrix  $W_{yh}$  if we are going to use softmax to classify? Ignore the bias unit here.**

- A. 500x1
- **B RIGHT.** 5x100
- C. 100x5
- D. 5x5

**10. What is the total number of parameters that we would have to train in case the RNN architecture used was LSTM. Ignore all bias units as well as the output parameters**

- A. 15000
- B. 15100
- C. 45000
- **D RIGHT.** 60000

LSTM have eight paramers; four **W** and four **U** (  $W_{xh}$  is the notation used above for U )

Size of W is (100,100) and U is (100, 50)

So Total  $4 \times 100 \times 100 + 4 \times 100 \times 50 = 60000$