Modern RNNs

Quiz, 4 questions

4/4 points (100%)

Congratulations! You passed!

Next Item

1 / 1
point

1.
Choose correct statements about the exploding gradient problem:

- [x] Exploding gradient problem is easy to detect.

Correct
Exploding gradients are easy to detect, not vanishing.

- [ ] ReLU nonlinearity helps with the exploding gradient problem.

Un-selected is correct

- [x] The reason of the exploding gradient problem in the simple RNN is the recurrent weight matrix W. Nonlinearities sigmoid, tanh, and ReLU does not cause the problem.

Correct
Derivatives of all these nonlinearities are less than 1, therefore they may cause only the vanishing gradient problem.

- [ ] The threshold for gradient clipping should be as low as possible to make the training more efficient.

Un-selected is correct

---

1 / 1
point

2.

Choose correct statements about the vanishing gradient problem:

Modern RNNs

Quiz, 4 questions

4/4 points (100%)

- [ ] Vanishing gradient problem is easy to detect.

Un-selected is correct

- [x] Both nonlinearity and the recurrent weight matrix W cause the vanishing gradient problem.

Correct
That is true!

- [x] Orthogonal initialization of the recurrent weight matrix helps with the vanishing gradient problem.

Correct
That is true!

- [ ] Truncated BPTT helps with the vanishing gradient problem.

Un-selected is correct

---

1 / 1
point

3.

Consider the LSTM architecture:

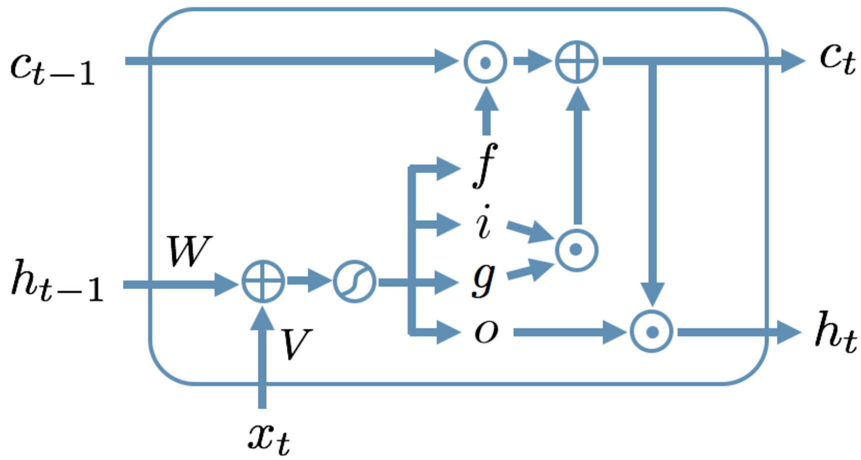
Modern RNNs

Quiz, 4 questions

4/4 points (100%)

$$\begin{pmatrix} g_t \\ i_t \\ o_t \\ f_t \end{pmatrix} = \begin{pmatrix} \tilde{f} \\ \sigma \\ \sigma \\ \sigma \end{pmatrix} (Vx_t + Wh_{t-1} + b)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot g_t, \quad h_t = o_t \cdot \tilde{f}\,(c_t)$$



Choose correct statements about this architecture:

- [x] The LSTM needs four times more parameters than the simple RNN.

Correct
For each gate we need its own set of parameters and there are 3 gates in the LSTM architecture.

- [ ] Gradients do not vanish on the way through memory cells c in the LSTM with forget gate.

Un-selected is correct

- [ ] There is a combination of the gates values which makes the LSTM completely equivalent to the simple RNN.

Un-selected is correct

- [x] The exploding gradient problem is still possible in LSTM on the way between h_{t-1} and h_t.

Correct
Very large norm of W may cause the exploding gradient problem. Therefore gradient clipping is useful for LSTM and GRU architectures too.

Modern RNNs

Quiz, 4 questions

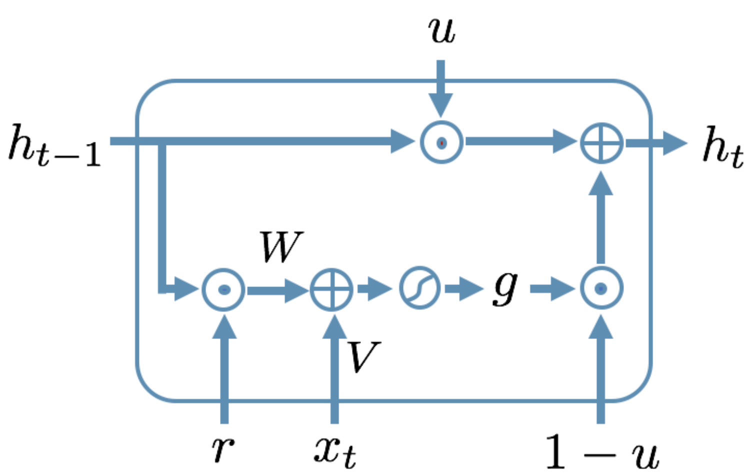
4/4 points (100%)

1 / 1
point

4.
Consider the GRU architecture:

$$g_t = \tilde{f}\left(V_g x_t + W_g(h_{t-1} \cdot r_t) + b_g\right)$$

$$h_t = (1 - u_t) \cdot g_t + u_t \cdot h_{t-1}$$



Which combination of the gate values makes this model equivalent to the simple RNN? Here value zero corresponds to a closed gate and value one corresponds to an open gate.

- ( ) Both reset and update gates are open.
- ( ) Both reset and update gates are closed.
- (•) Reset gate is open and update gate is closed.

Correct
That is it!

- ( ) Update gate is open and reset gate is closed.

Modern RNNs

Quiz, 4 questions

4/4 points (100%)