Introduction to Deep Learning (CS474)

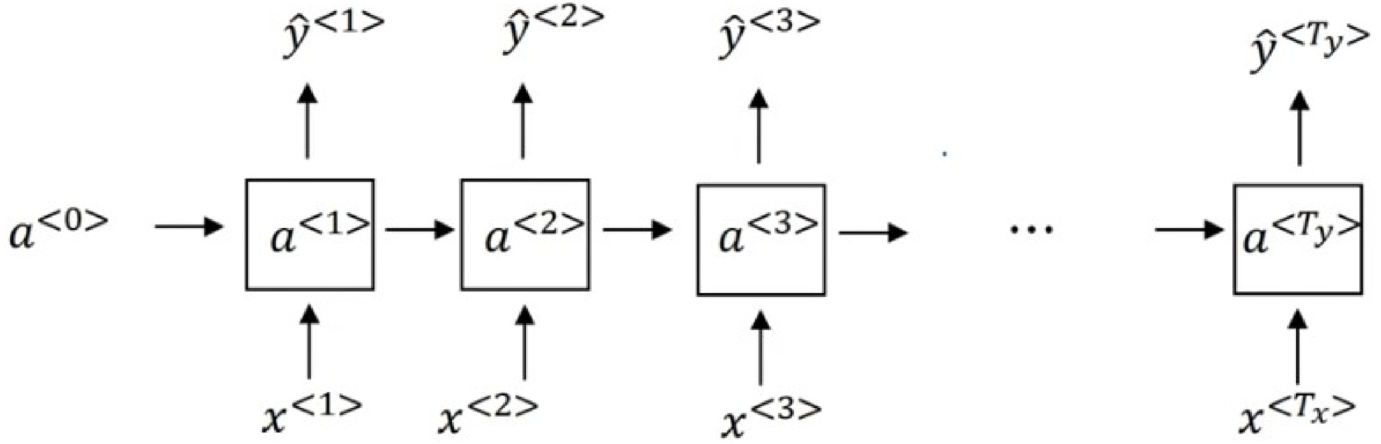
Lecture 24

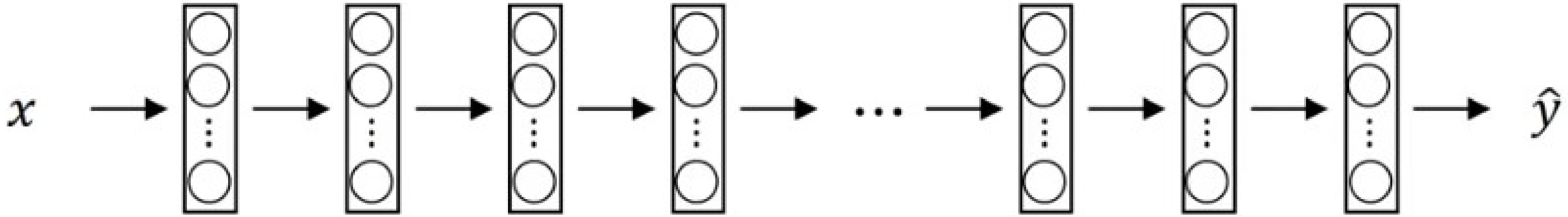
Rupam

Outline

- **Module 3**
  - **Problems with RNN**
  - **Gated Recurrent Unit**

Vanishing Gradients





Slide credit: Prof. A. Ng

Introduction to Gated Recurrent Units (GRU)

- We have introduced the basics of RNNs, which can better handle sequence data.
- For demonstration, we implemented RNN using PyTorch for solving a simple problem.
- However, such techniques may not be sufficient for practitioners when they face a wide range of sequence learning problems nowadays.

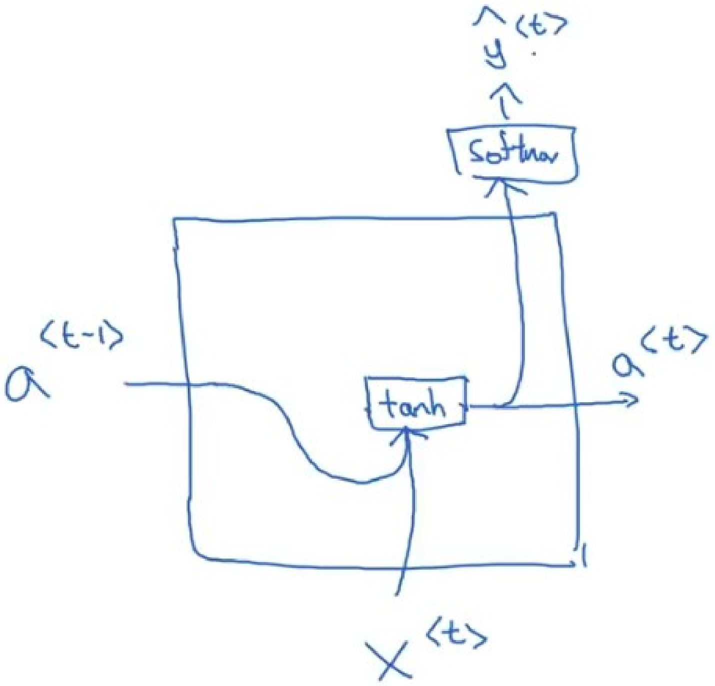
Slide credit: A. Zhang, Z. C. Lipton, M. Li, and A. Smola

Introduction to Gated Recurrent Units (GRU)

- The key distinction between vanilla RNNs and GRUs is that the latter support gating of the hidden state.
- This means that we have dedicated mechanisms for when a hidden state should be updated and also when it should be reset.
- For instance, if the first token is of great importance we will learn not to update the hidden state after the first observation.
- Likewise, we will learn to skip irrelevant temporary observations.
- Last, we will learn to reset the latent state whenever needed.

Slide credit: A. Zhang, Z. C. Lipton, M. Li, and A. Smola

Recap!



$$a^{<t>} = g(W_a[a^{<t-1>}, x^{<t>}] + b_a)$$

Slide credit: Prof. A. Ng

GRU (Simplified)

c = memory cell

$$c^{<t>} = a^{<t>}$$

$$\tilde{c}^{<t>} = \tanh(W_c[c^{<t-1>}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[c^{<t-1>}, x^{<t>}] + b_u)$$

"update"

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1-\Gamma_u) * c^{<t-1>}$$

Slide credit: Prof. A. Ng

Rupam

GRU (Simplified)

$$\tilde{c}^{<t>} = \tanh(W_c[\,c^{<t-1>}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[\,c^{<t-1>}, x^{<t>}] + b_u)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) * c^{<t-1>}$$

Slide credit: Prof. A. Ng

GRU

$$\underset{\sim}{\tilde{c}^{<t>}} = \tanh(W_c[\,\Gamma_r * \underset{\sim\sim}{c^{<t-1>}}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[\,c^{<t-1>}, x^{<t>}] + b_u)$$

$$\Gamma_r = \sigma(W_r[c^{<t-1>}, x^{<t>}] + b_r)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) * c^{<t-1>}$$

Slide credit: Prof. A. Ng

Rupam

Summary

In summary, GRUs have the following features:

- Gated RNNs can better capture dependencies for sequences with large time step distances.
- Reset gates help capture short-term dependencies in sequences.
- Update gates help capture long-term dependencies in sequences.

Slide credit: A. Zhang, Z. C. Lipton, M. Li, and A. Smola

References

- All the contents present in the slides are taken from various online resources. Due credit is given in the respective slides.These slides are used for *academic* purposes only.