Hide menu

LSTMs and Named Entity Recognition

- Video: Week Introduction 1 min
- Video: RNNs and Vanishing Gradients 6 min
- Reading: RNNs and Vanishing Gradients 6 min
- Reading: (Optional) Intro to optimization in deep learning: Gradient Descent 10 min
- Lab: Vanishing Gradients 15 min
- Video: Introduction to LSTMs 4 min
- Reading: Introduction to LSTMs 3 min
- Video: LSTM Architecture 3 min
- Reading: LSTM Architecture 4 min
- Video: Introduction to Named Entity Recognition 3 min
- Reading: Introduction to Named Entity Recognition 2 min
- Video: Training NERs: Data Processing 4 min
- Reading: Training NERs: Data Processing 5 min
- Reading: Long Short-Term Memory (Deep Learning Specialization C5) 10 min
- Video: Computing Accuracy 1 min
- Reading: Computing Accuracy 2 min
- Video: Week Conclusion 32 sec

Lecture Notes (Optional)

Practice Quiz

Assignment: Named Entity Recognition (NER)

Week 2 > RNNs and Vanishing Gradients

Previous Next

RNNs and Vanishing Gradients

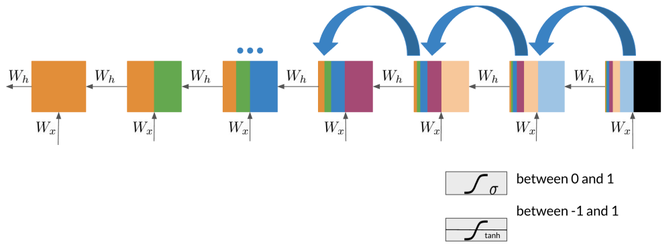
Advantages of RNNs

RNNs allow us to capture dependencies within a short range and they take up less RAM than other n-gram models.

Disadvantages of RNNs

RNNs struggle with longer term dependencies and are very prone to vanishing or exploding gradients.

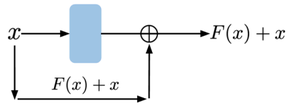
Note that as you are back-propagating through time, you end up getting the following:



Note that the *sigmoid* and *tanh* functions are bounded by 0 and 1 and -1 and 1 respectively. This eventually leads us to a problem. If you have many numbers that are less than |1|, then as you go through many layers, and you take the product of those numbers, you eventually end up getting a gradient that is very close to 0. This introduces the problem of vanishing gradients.

Solutions to Vanishing Gradient Problems

- Identity RNN with ReLU activation $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ -1 ⟶ 0
- Gradient clipping 32 ⟶ 25
- Skip connections



Mark as completed