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

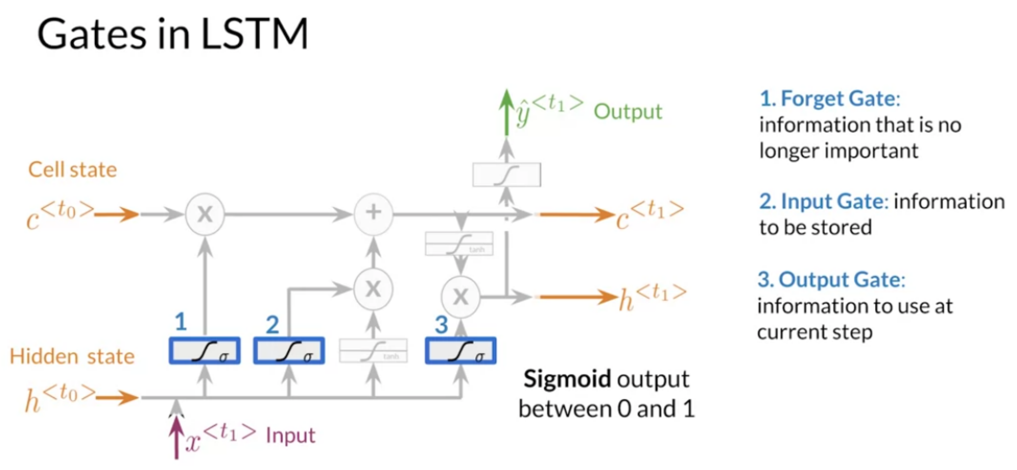
Lecture Notes (Optional)

Practice Quiz

Assignment: Named Entity Recognition (NER)

LSTM Architecture

The LSTM architecture could get complicated and don't worry about it if you do not understand it. I personally prefer looking at the equation, but I will try to give you a visualization for now and later this week we will take a look at the equations.



Gates in LSTM

1. Forget Gate: information that is no longer important

2. Input Gate: information to be stored

3. Output Gate: information to use at current step

Sigmoid output between 0 and 1

Note the forget gate (1), input gate (2) and output gate (3) marked in blue. In contrast with vanilla RNNs, there is the cell state in addition to the hidden state. The idea of the **forget gate** to drop the information that is no longer important. It makes use of the previous hidden state $h^{<t0>}$ and the input $x^{<t1>}$. The **input gate** makes sure to keep the relevant information that needs to be stored. Finally the **output gate** creates an output that is used at the current step.

LSTM equations (optional):
For better understanding, take a look at the LSTM equations and relate them to the figure above.

The forget gate: $f = \sigma\left(Wf\ [ht{-}1; xt] + bf\right)$ (marked with a blue 1)

The input gate: $i = \sigma\left(Wi\ [ht{-}1; xt] + bi\right)$ (marked with a blue 2)

The gate gate (candidate memory cell): $g = \tanh\left(Wg\ [ht{-}1; xt] + bg\right)$

The cell state: $ct = f \odot ct{-}1 + i \odot g$

The output gate: $o = \sigma\left(Wo\ [ht{-}1; xt] + bo\right)$ (marked with a blue 3)

The output of LSTM unit: $ht = ot \odot tanh(ct)$

Mark as completed

Like Dislike Report an issue

coach