Recurrent Nets for Sequence Data
Why Recurrent Structure

- Solve problems of sequence data: speeches, languages
- Captures sequence dependencies
Recurrent Structure: Folded and Unrolled View

Folded View

Unrolled View

outputs

states

inputs

$Y_t$

$s_t$

$x_t$

$Y_1$

$s_1$

$x_1$

$Y_2$

$s_2$

$x_2$

$Y_3$

$s_3$

$x_3$
Simple Recurrent Net

transition

\[ s_t = \sigma(W_1 s_{t-1} + W_2 x_t) \]

emission

\[ y_t = \sigma(W_3 s_t) \]

Hard to capture long term dependency, exponential multiplication effect
LSTM: Update with Moving Average

\[ s_t = \alpha_t s_{t-1} + \beta_t \tanh(W[x_t, s_{t-1}]) \]

How to set the forget rate and remembering rate?

Use another neural net module
One variant of LSTM

\[ s_t = \alpha_t s_{t-1} + \beta_t \tanh(W[x_t, s_{t-1}]) \]
\[ \alpha_t = \sigma(W_f[x_t, s_{t-1}]) \]
\[ \beta_t = \sigma(W_r[x_t, s_{t-1}]) \]

Many other variants exist with similar spirit.
Unrolled Stacked Recurrent Nets

- Provide Hierarchical representation about sequence
- Feed output of one sequence to another RNN

![Diagram of Unrolled Stacked Recurrent Nets]
LSTM as Compositional Building Block

**Sequence Encoder**
sequence to vector

**Sequence Decoder**
vector to sequence
Seq2Seq: Machine Translation Model

source sequence

target sequence
Image2Seq: Image Caption Model

Image

ConvNet

Decoder

dog

catch

ball
It is all about Composability!
Topics not Covered Today

● More complicated memory structures
  ○ Memory networks

● Objective derivation for generative models
  ○ Generative adversarial nets
  ○ Variational methods

● Gradient estimation with hard decision and interactions
  ○ Q learning, Policy gradient