RECURRENT NEURAL NETWORKS (RNNs)

1. Motivation
   (a) Sequences \(<x_1, x_2, ..., x_T>, <y_1, ..., y_T>\) are everywhere!

   → How do we model input/hidden/output sequences?

   [Especially when sequence length is not fixed/constant]

2. Most General Formulation

   → No inputs

   \[ h_t = f_0(h_{t-1}) \]

   Think Markov Chain!

   \( f = \text{non-linear function} \quad \Theta = \text{parameters} \)
→ Sequential input

\[ h_t = f_0(x_t, h_{t-1}) \]

Think HMMs!

→ Sequential input & output

\[ h_t = f_0(x_t, h_{t-1}) \]

\[ y_t = g_0(h_t) \]

Think Kalman Filtering!
3. RNNs: aka How about we implement $f_{o_c}$ with neural nets.

Recall: Multi-Layer Perceptrons (MLPs)

Layer $l$

$\overrightarrow{\mathbf{h}}_l = \mathbf{h}^{(l)}_0$, $\mathbf{b}^{(l)}$

Layer $2$

$\overrightarrow{\mathbf{h}}_2 = \mathbf{h}^{(2)}_0$, $\mathbf{b}^{(2)}$

Layer $t$

$\overrightarrow{\mathbf{h}}_t = \mathbf{h}^{(t)}_0$, $\mathbf{b}^{(t)}$

Parameter $\mathbf{w}^{(g)}$, $\mathbf{b}^{(g)}$

$\mathbf{z} = \mathbf{h}^{(g)}_t$

$\mathbf{h}^g_l = \sigma (\mathbf{w}_l^{(g)} \mathbf{h}^{(l)}_t + \mathbf{b}^{(g)})$

some non-linearity (tanh, sigmoid, ReLU)

RNNs: How about we add feedback loops

Each layer shares parameters ($\mathbf{w}^{(g)}$, $\mathbf{b}^{(g)}$)
Think of a single "cell" or "step" (t+1)

\[ h^l_t \rightarrow h^l_{t-1} \rightarrow h^l_t \]

In principle, these could be different

→ Key "Forward-Pass" Computation

\[ h^l_t = f_0(h^l_{t-1}, h^l_{t-1}) \]

→ "Vanilla" RNN

\[ \hat{h}^l_t = \sigma \left( W^{l-1} h^l_{t-1} + b^l \right) \]  -- (i)

(element-wise tanh)

→ Note: Sometimes:

\[ h^l_t = W^{l-1} \sigma(h^l_{t-1} \cdot \sigma(h^l_{t-1})) + b^l \]  -- (ii)
Claim: The two forms (i) & (ii) are equivalent.

Why?

Proof Sketch: Just different ways of incorporating non-linearity. Either at start or end.

\[ h_0 \xrightarrow{\text{Inner Product}} \xrightarrow{\text{Non-linearity}} \xrightarrow{\text{IP}} \xrightarrow{\text{Non-linearity}} \]

Approach (i) calls this a layer

Approach (ii) calls this a layer

\[ \text{So that's a "Vanilla" RNN.} \]

Generalizations:

- change non-linearity (ReLU)
- change cells altogether (LSTM, GRU)

- Change loops / graph structure
  e.g. what if layer 3 looped back to layer 2?

\[ \text{What do we get?} \]

[Hint: not a grid]