CS 152—Notes on Gated Units

Neil Rhodes

December 2, 2019

1 Gated Units

Gated units provide a mechanism for storing data in an RNN for long periods of time.

1.1 Long Short Term Memory—LSTM

An LSTM unit adds an additional value to a standard RNN unit. Rather than just having the hidden value: h_t, there is also a stored cell c_t. There are three gates: the forget gate, f, the input gate, i, and the output gate, o. The forget gate controls whether the stored cell is copied from the previous timestep. The input gate controls whether input from the current x and previous h are input into c, Finally, the output gate controls whether c is output into the current h.

The equations are:

$$f_t = \sigma(x_t W_{xf} + h_{t-1} W_{hf} + b_f) \tag{1}$$
$$i_t = \sigma(x_t W_{xi} + h_{t-1} W_{hi} + b_i) \tag{2}$$
$$o_t = \sigma(x_t W_{xo} + h_{t-1} W_{ho} + b_o) \tag{3}$$

Now, we update c_t and h_t:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(x_t W_{xc} + h_{t-1} W_{hc} + b_c) \tag{4}$$
$$h_t = o_t \odot \tanh(c_t) \tag{5}$$

1.2 Gated Recurrent Unit—GRU

GRU cells store just the hidden value: h_t. There are two gates: the reset gate: r_t, and the update gate: z_t.

The equations are:

$$r_t = \sigma(x_t W_{xr} + h_{t-1} W_{hr} + b_r) \tag{6}$$
$$z_t = \sigma(x_t W_{xz} + h_{t-1} W_{hz} + b_z) \tag{7}$$
$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tanh(x_t W_{xh} + (r_t \odot h_{t-1}) W_{hh} + b_h) \tag{8}$$