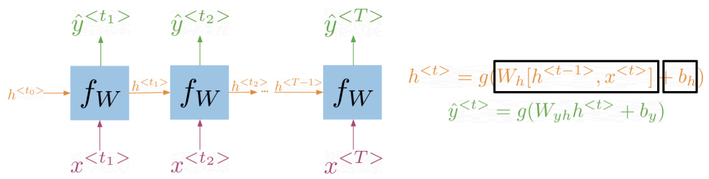
Math in Simple RNNs

It is best to explain the math behind a simple RNN with a diagram:



$$h^{<t>} = g(W_h[h^{<t-1>}, x^{<t>}] + b_h)$$

$$\hat{y}^{<t>} = g(W_{yh}h^{<t>} + b_y)$$

$$h^{<t>} = g(W_{hh}h^{<t-1>} \oplus W_{hx}x^{<t>} + b_h)$$

Note that:

$$h^{<t>} = g(Wh[h^{<t-1>}, x^{<t>}] + bh)$$

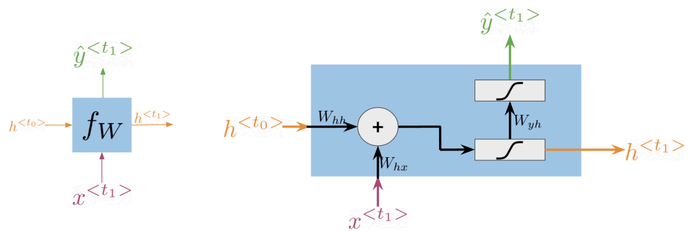
Is the same as multiplying Whh by h and Whx by x. In other words, you can concatenate it as follows:

$$h^{<t>} = g\left(Whhh^{<t-1>} \oplus Whxx^{<t>} + bh\right)$$

For the prediction at each time step, you can use the following:

$$\hat{y}^{<t>} = g\left(Wyhh^{<t>} + by\right)$$

Note that you end up training Whh, Whx, Wyh, bh, by. Here is a visualization of the model.



Mark as completed

Like Dislike Report an issue