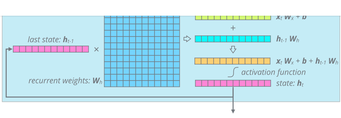
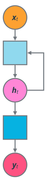


Recurrent Layer: Internals

- The input word vector x_t is first multiplied by the weight matrix W_x.
- Then bias values are added to produce our first intermediate result: $x_t W_x + b$
- Meanwhile, the state vector from the previous time step h_{t-1} is multiplied with another weight matrix W_h to produce our second intermediate result: $h_{t-1} W_h$.
- These two are then added together, and passed through an activation function such as ReLU, sigmoid or tanh to produce the state for the current time step: h_t.
- This state vector is passed on as input to the next fully-connected layer, that applies another weight matrix, bias and activation to produce the output: y_t

Let's simplify this diagram and look at the bigger picture again.



Basic RNN: Schematic

The key thing to note here is that the RNN's state h_t is used to produce the output y_t, as well as looped back to produce the next state.

In summary, a recurrent layer computes the current state h_t as:

$$h_t = f(x_t W_x + h_{t-1} W_h + b)$$

Here $f(\cdot)$ is some non-linear activation function, x_t is the input vector, W_x is the input weight matrix, h_{t-1} is the previous state vector, W_h is the recurrent weight matrix and b is the bias vector.

Quiz Question

Let's say you've decided to use a word vector (x_t) of length 200, and a state vector (h_t) of length 300. Treating these as single-row matrices, we can write the sizes as 1x200 and 1x300 respectively.

Now, what is the size of each parameter of the RNN? Match the correct sizes below.

200x300 | 200x200 | 300x1 | 1x300

Parameter	Size (rows x cols)
Input weight matrix (W_x)	200x300
Recurrent weight matrix (W_h)	300x300
Bias vector (b)	1x300

Submit