from z_{t-1} it received, only to the degree the gates of the node opted to let information through to change it, making it easy for information to proceed along unchanged between iterations.

Three of these gates update the hidden state z of the LSTM at each iteration, and filter what part will constitute output y:

The *forget gate* layer (in red): the first decision is what part of the data of the hidden state z_{t-1} to discard. The forget gate sigmoid layer takes y_{t-1} and x_t and produces a number between 0 and 1 for each value of the hidden state z_{t-1}, dictating whether to completely discard, partially discard or keep that data.

The *input gate* layer (in green): the next decision is what new data should be added to the hidden state. This is comprised of two parallel elements: the input gate sigmoid layer sets which values will be updated, and in parallel the input gate *tanh* layer creates a vector of new candidate values $\hat{z}_t$ that might be added. When the vector of candidate values $\hat{z}_t$ is multiplied by the sigmoid output f_t , the product is that part of the candidate values that will be let through to be added to the hidden state.

Once both the forget gate and the input gate have updated the hidden state z_{t-1}, z_t is ready: $z_t = f_t * z_{t-1} + i_t * \hat{z}_t$

The *output gate* layer (violet): The final decision is what parts of the new hidden state z_t to produce as output. The output gate sigmoid layer acts as a filter setting what will be output and what will remain "hidden". The hidden state is run through *tanh* (to normalize values between -1 and 1) and then multiplied by the sigmoid layer output: $y_t = o_t * tanh(z_t)$

Now that we've understood the basic structure of an RNN/LSTM unit, let's look at using these units to form a network that will be able to perform sequence transduction, also known as: **sequence-to-sequence**.

(C) ENCODER-DECODER:

Neural sequence transduction models have an encoder-decoder structure. The encoder transforms an input sequence x to a hidden sequence z. Given z, the decoder can produce an output sequence y.

The encoder LSTM produces no external output, only the hidden state z is updated at each iteration. Once the encoder has completed producing z_n , the decoder takes that as its initial input z'_0.