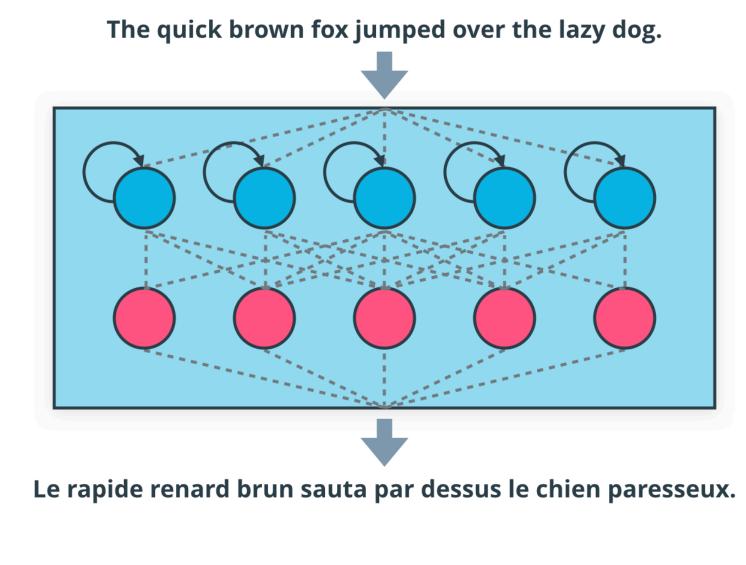
UDACITY My Programs Catalog

Machine Translation

Machine Translation can be thought of as a sequence-to-sequence learning problem.



You have one sequence going in, i.e. a sentence in the source language, and one sequence coming out, its translation in the target language.

Machine Translation: A Sequence-to-Sequence Learning Problem

either use a dense or fully-connected layer to produce the output, or use another decoding layer.

decoding layer.

Experimenting with different network architectures and recurrent layer units (such as

LSTMs, GRUs, etc.), you can come up with a fairly simple model that performs decently well on small-to-medium size datasets. Commercial-grade translation systems need to deal with a much larger vocabulary, and hence have to use a much more complex model, apply different optimizations, etc. Training such models requires a lot of data and compute time.

This seems like a very hard problem - and it is! But recent advances in Recurrent Neural Networks have shown a lot of improvement. A typical approach is to use a recurrent layer to encode the meaning of the sentence by processing the words in a sequence, and then

Neural Net Architecture for Machine Translation

Let's develop a basic neural network architecture for machine translation.

Input Representation

The key thing to note here is that instead of a single word vector or document vector as

input, we need to represent each sentence in the source language as a sequence of word vectors. Therefore, we convert each word or token into a one-hot encoded vector, and

stack those vectors into a matrix - this becomes our input to the neural network.

quick

brown

the one-hot encoded input word vectors

fox **X**3 jumped **X**4 over **X**5 the **X**6 lazy **X**7 dog **X**8 Input Representation You may be wondering what to do about sequences of different length: One common approach is to simply take the sequence of maximum length in your corpus, and pad each sequence with a special token to make them all the same length.

X0

X1

X2