Review Neural Sequence Models

Besides images, another common data modality is text. Simple multilayer Perceptrons are not ideal for text data for two reasons:

1. Text data consists of words, not vectors, and Perceptrons take in vectors as input. One can compute bag-of-word vectors, but these are very high dimensional and you lose any information conveyed in the word order.
2. Text data is inherently sequential and context-dependent. Often, a word itself is not meaningful until put into a sequence. For instance, consider the following sentences:
   - He went to the bank to withdraw some money.
   - He likes to go on a stroll at the river bank after dinner.

☆ **Key Points**

There are two components to sequence models: encoders and decoders.

Encoders take in a sequence of words and outputs a vector (or a code) that can be viewed as a summary of the input sequence.

Decoders take in such a vector (computed from encoder) and turn it into a scalar or sequence of outputs.

If we only look at the single word “bank,” it is hard to tell whether the bank we are referring to is a financial bank or a river bank. On the other hand, the meaning is clear if we look at the whole sentence (or the whole sequence of words).

Special types of neural networks have been developed to deal with text data. In the beginning of this module, we will explore word embeddings, a technique that AI researchers have come up with to turn words into vectors. Next, we will talk about neural sequence models that are used to handle different text applications such as machine translation, text summarization, question answering, and so on. These models usually consist of two components:

1. **Encoder:** A neural network that takes in a sequence of words (represented using word embeddings) and outputs a vector or a code that can be viewed as a summary of the input sequence.
2. **Decoder:** A neural network that takes in the vector output of an encoder and turns it into a scalar or sequence of outputs. (These can be words represented by word embeddings or other things, depending on the application.)

Typically, the encoder and decoder are learned jointly. For example, in neural machine translation, one trains on input, output translation pairs. For example, these could be English, German sentence pairs and one trains the encoder and decoder to map the English sentence to its German translation.