Recurrent Neural Networks (RNNs)

Have you ever watched a TV series or movie where the story builds up episode by episode, and each episode depends on the previous one? That's exactly how Recurrent Neural Networks work!

RNNs are a type of neural network designed to handle sequential data, where the order of the data is important. They can process input of varying lengths and remember information from the previous inputs to make predictions about the current input.

To understand this better, let's take an example of a language model. Suppose we want to train a machine to generate a sentence given a few starting words. For this, we can use an RNN where each word in the sentence is considered a time step.

During training, the RNN takes in each word as input and uses it to make a prediction about the next word in the sequence. It then uses the predicted word and the previous hidden state to make a prediction about the next word, and so on.

At the end of each time step, the RNN updates its hidden state with the information from the previous time step, allowing it to remember information from earlier in the sequence. This is called "recurrence" because the network is reusing the same weights for each time step.

RNNs are useful in many real-life applications, such as speech recognition, natural language processing, and image captioning. They can be used to predict stock prices, generate music, or even to create a chatbot that can respond to messages in a conversational manner.

Overall, RNNs are a powerful tool for handling sequential data, and with the right training and parameters, they can make accurate predictions about the future based on the past.