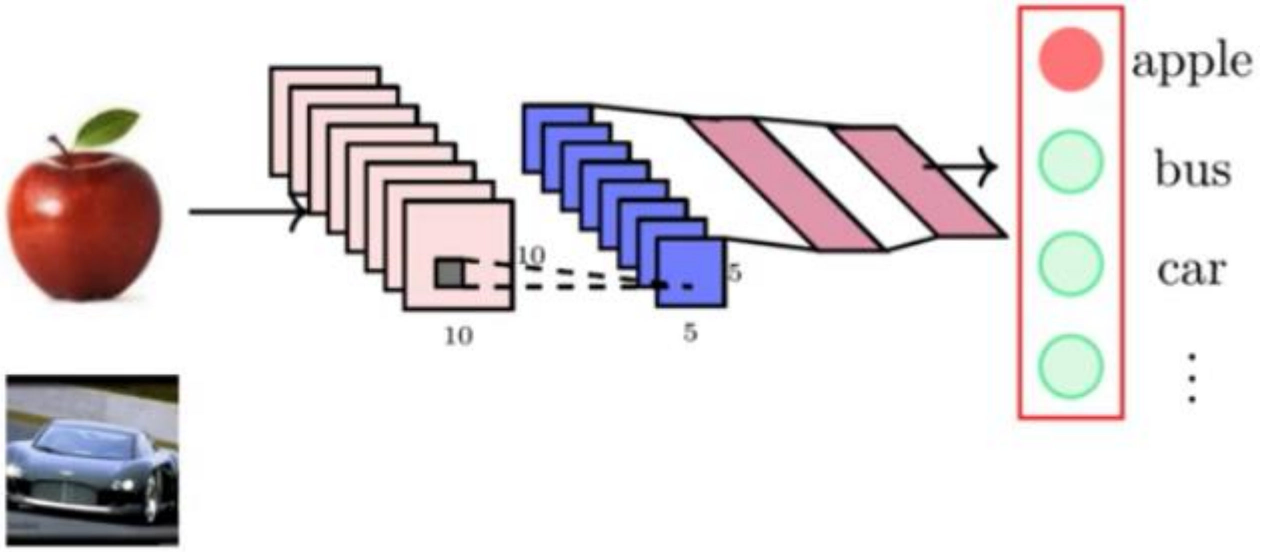
Module 6

Introduction to Sequence Modeling Problems

Consider the problem of predicting the health risk of a person based on multiple health parameters and we have decided to model the true relationship between the input and output using Feed-forward neural networks (also known as **Multi-layered Network of Neurons**).

In Feed-forward Neural Networks (FNN) the output of one data point is completely independent of the previous input i.e… the health risk of the second person is not dependent on the health risk of the first person and so on.

Similarly, in the case of Convolution Neural Networks (CNN), the output from the softmax layer in the context of image classification is entirely independent of the previous input image.



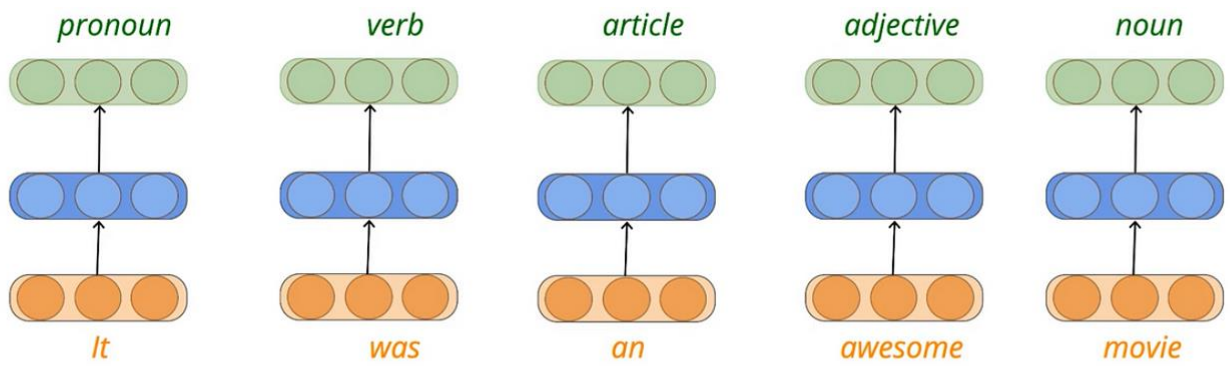
The characteristics of FNN and CNN's are:

Let's look at the problem of auto-complete in the context of sequence modeling. In this problem, whenever we type a character (**d**) the system tries to predict the next possible character based on the previously typed character.

In other words, the network tries to predict the next character from the possible 26 English alphabets given that we have typed '**d**'. The neural network would have a softmax output of size 26 representing the probability of the next letter given the previous letters. Since the inputs to this network are characters we need to convert them to a one-hot encoded vector of size 26 and the element corresponding to the index of the alphabet would be set to 1, everything else is set to 0.

Module 6

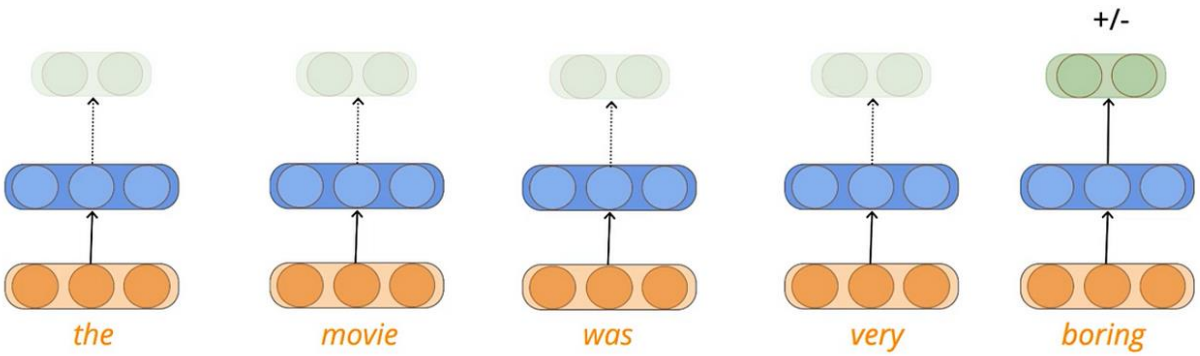
Parts of Speech Tagging



In the problem of parts of speech tagging, we are given a sequence of words for every word we need to predict the part of speech tag for that word (eg: verb, noun, pronoun, etc…). Again in this problem, the output is not only dependent on the current input (current word) but also on the previous input. For example, the probability of tagging the word 'movie' as a **noun** would be higher if we know that the previous word is an adjective.

Sequence Classification

Do we need to produce output at every time step?

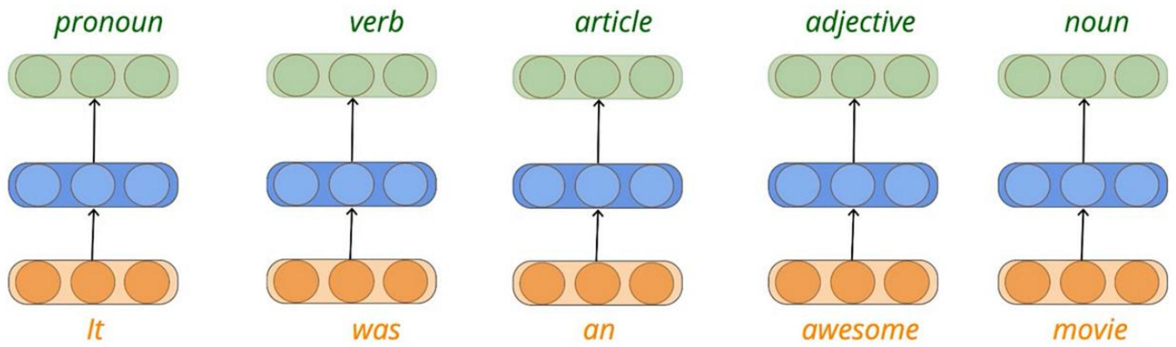


Imagine that you want to predict the pulse of the movie by analyzing the reviews. In this scenario, we don't need to output after every word of the input rather we just need to understand the mood after reading the entire sentence i.e…either positive or negative. Understanding the mood from a text with machine learning is called **Sentiment analysis**.

Modeling Sequence Learning Problems

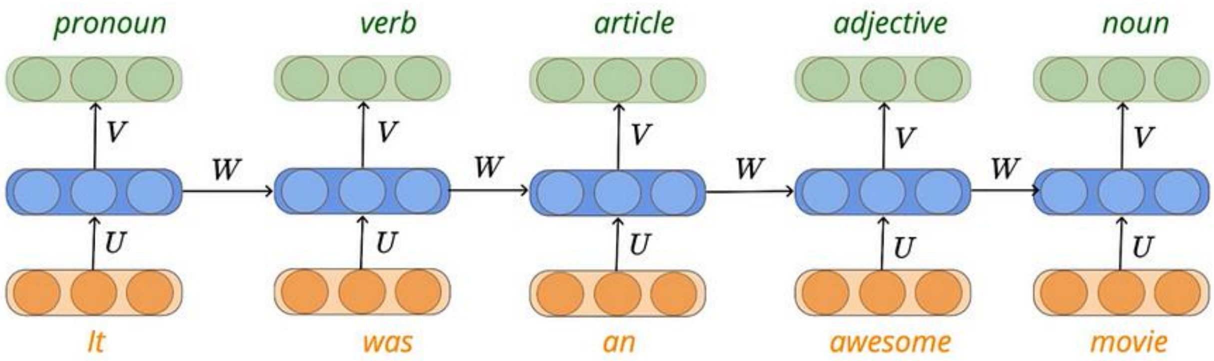
In the previous section, we have learned about practical applications of sequence learning problems but how we model such problems?.

Module 6



Recurrent Neural Networks

Recurrent Neural Networks(RNN) are a type of Neural Network where the output from the previous step is fed as input to the current step.



In RNN, you can see that the output of the first time step is fed as input along with the original input to the next time step.

The input to the function is denoted in orange color and represented as an $\mathbf{x_i}$. The weights associated with the input is denoted using a vector **U** and the hidden representation (**s_i**) of the word is computed as a function of the output of the previous time step and current input along with bias. The output of the hidden represented (**s_i**) is given by the following equation,

$$s_i = \sigma(Ux_i + Ws_{i-1} + b)$$
$$y_i = O(Vs_i + c)$$

Once we compute the hidden representation of the input, the final output ($\mathbf{y_i}$) from the network is a softmax function of hidden representation and weights associated with it along with the bias. We

Module 6

are able to come up with an approximate function that is able to satisfy all the three conditions that we have set to solve the problems of sequence learning.

$$y_i = \hat{f}(x_i, s_{i-1}, W, U, V, b, c)$$