The Maths involved in Voice Recognition

An ML Assignment

Nikhil P

Indian Institute of technology
Hyderabad

January 21, 2020

Table of contents

- Zero Padding
- MFCC
- Recurrring Neural Network
- Loss Function

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Zero Padding

Zero padding consists of extending a signal (or spectrum) with zeros. It maps a length N signal to a length $M > N$ signal, but N need not divide M. This is used to multiply the original 80 files to 2000 in number.

$$\textsc{ZeroPad}_{M,m}(x) \triangleq \begin{cases} x(m), & |m| < N/2 \\ 0, & \text{otherwise} \end{cases}$$

where $m = 0, \pm 1, \pm 2, \ldots, \pm M_h$, with $M_h \triangleq (M-1)/2$ for M odd, and $M/2 - 1$ for M even.

example,

$$\textsc{ZeroPad}_{10}([1,2,3,4,5]) = [1,2,3,0,0,0,0,0,4,5].$$

Code: 250files.py

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

MFCC

The features of the given audio sample is detected by first changing the clip to a matrix and using the MFCC function from Librosa. This gives a matrix of dimension 49x39 i.e the file is broken down to 49 time-steps, with each step having 39 features.

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Recurrent Neural Network

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

In traditional neural networks, all the inputs and outputs are independent of each other, but in cases like when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words.

Thus RNN came into existence, which solved this issue with the help of a Hidden Layer.

The main and most important feature of RNN is Hidden state, which remembers some information about a sequence.

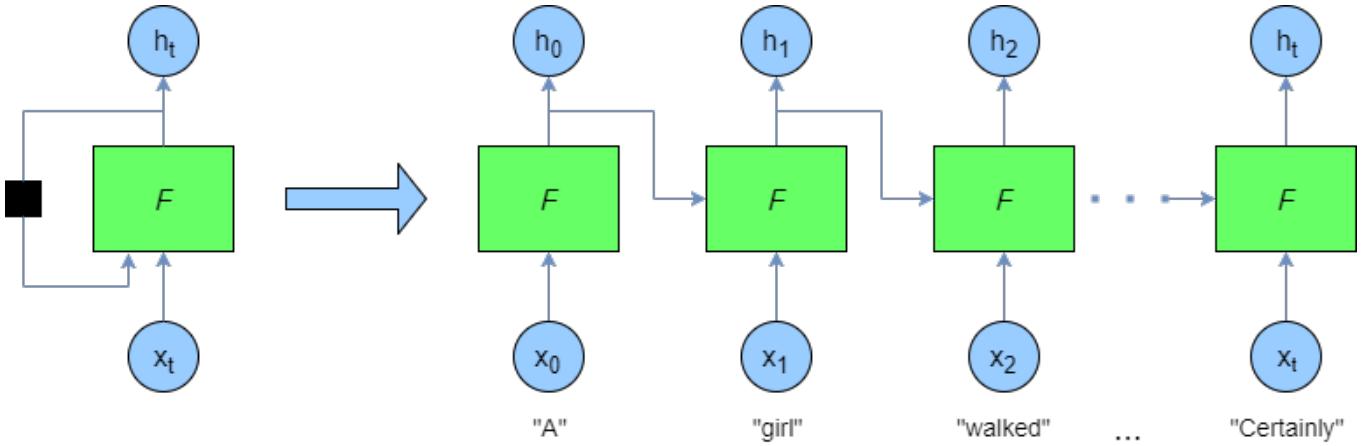
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

In a normal Nueral Network, each hidden layer will have its own set of weights and biases, let's say, for hidden layer 1 the weights and biases are (w1, b1), (w2, b2) for second hidden layer and (w3, b3) for third hidden layer.
This means that each of these layers are independent of each other, i.e. they do not memorize the previous outputs.

Now, RNN does the following

(*A*) RNN converts the independent activations into dependent activations by providing the same weights and biases to all the layers, thus reducing the complexity of increasing parameters and memorizing each previous outputs by giving each output as input to the next hidden layer.

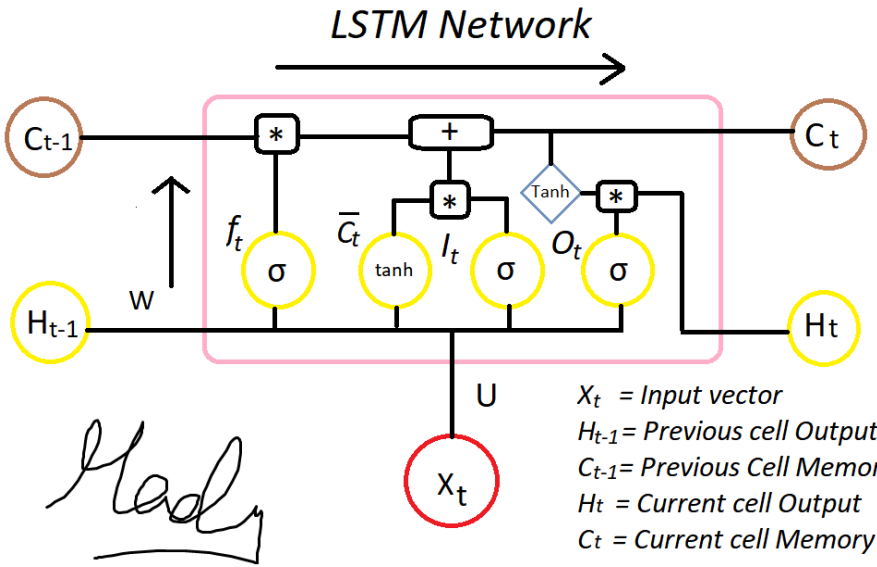
(*B*) Hence these three layers can be joined together such that the weights and bias of all the hidden layers is the same, into a single recurrent layer.



भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

LSTM

Here,we use a variant of RNN,called LSTM (Long Short Term Memory), in which a gate called forget gate is implemented, that makes the output from non-related nodes insignificant, i.e the overall weight given depends only on the relevant nodes.



X_t = Input vector
H_{t-1} = Previous cell Output
C_{t-1} = Previous Cell Memory
H_t = Current cell Output
C_t = Current cell Memory

(*) = Element-wise multiplication
(+) = Element-wise addition

$$f_t = \sigma(X_t * U_f + H_{t-1} * W_f)$$
$$\bar{C}_t = \tanh(X_t * U_c + H_{t-1} * W_c)$$
$$I_t = \sigma(X_t * U_i + H_{t-1} * W_i)$$
$$O_t = \sigma(X_t * U_o + H_{t-1} * W_o)$$

$$C_t = f_t * C_{t-1} + I_t * \bar{C}_t$$
$$H_t = O_t * \tanh(C_t)$$

W, U = weight vectors for forget gate (f), candidate (c) , i/p gate (I) and o/p gate (O)

Note : These are different weights for different gates, for simpicity's sake, I mentioned W and U

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Loss Function

The output of the LSTM is compared to the required output during training by the use of a loss function called Categorical Cross Entropy given by

$$E = -\sum_{i=0}^{C} y_i log(\hat{y}_i),$$

,where C is the total no. of Classes

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad