Recurent Neural Network

Tadipatri Uday Kiran Reddy
EE19BTECH11038
Dept. of Electrical Engg.,
IIT Hyderabad.

January 22, 2020

1. Zero Padding
2. Feature Extraction
3. Recurrent Neural Networks
4. LSTM

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Zero Padding

This is a technique to increase no of samples of a given audio sample

- Declare an empty array of fixed length which is same length for training the data
- Add the data of the audio sample in the array
- Do this in a circular shift way until desired no of samples are required

MFCC - Mel Frequency Cepstral Coefficient

MFCC takes into account human perception for sensitivity at appropriate frequencies by converting the conventional frequency to Mel Scale.

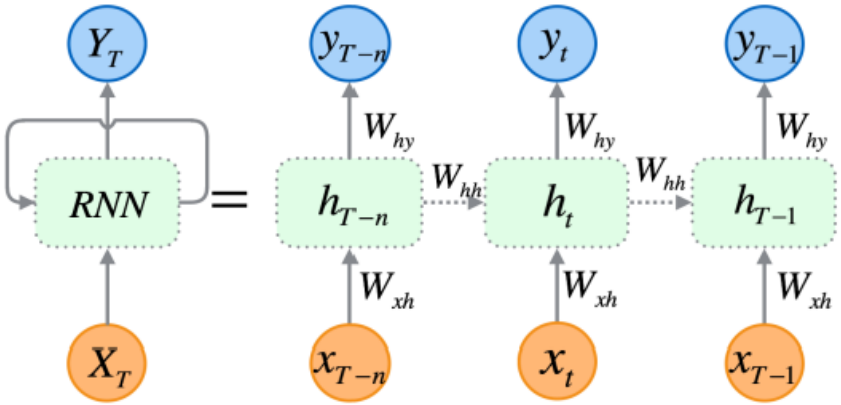
$$Mel(f) = 1125 \ln(1 + \frac{f}{700})$$

In our case, it returns a matrix of (49X39) i.e,49 time steps and each with 39 features.

Motivation for Recurrent Neural Network

The main advantage of the neural neural network over others is its memory for processing Temporal signals
Where as in Feed forward Neural Networks such as MLPN's at time t for a input $x(\mathrm{f})_{t\geq 0}$
output would be $\mathrm{f}(\mathrm{w},\mathrm{x}(\mathrm{t}))_{t\geq 0}$, **without regard to the previous history**.

Recurrent Neural Networks



1 Standard RNN architecture and an unfolded structure with T time

Here h_t is the hidden state at time step t

$$h_t = f(x_t W_{xh} + W_{hh} h_{t-1})$$

where f is a non-linear function

For initial hidden state is initialised with all zeros

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Back Propagation

$$o_t = softmax(V_{st})$$

$$softmax(\mathbf{z})_i = \sigma(\mathbf{z}) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}};$$

for i = 1,2,3,,,,K and z = $(z_1, z_2, z_3, , , , z_K)$

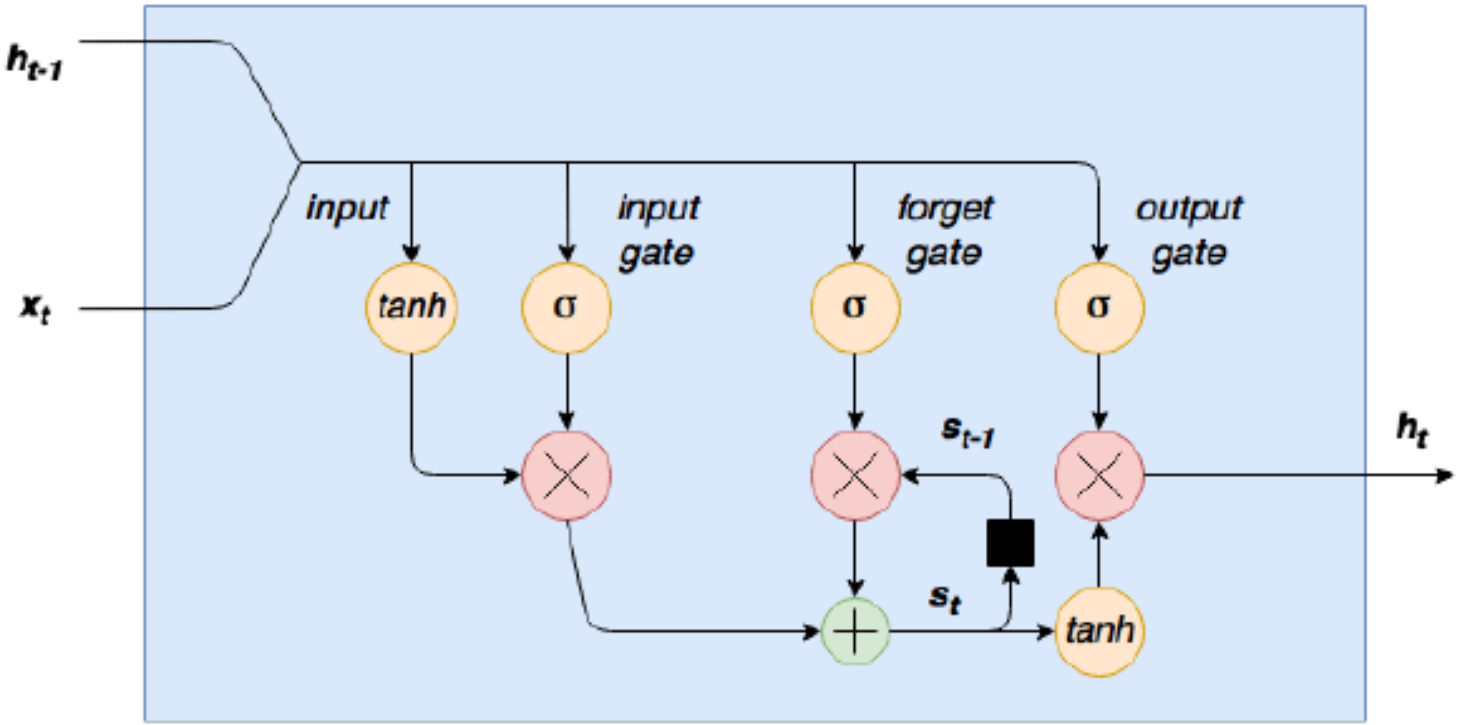
$$E(y_t, \hat{y}_t) = -y_t \log \hat{y}_t$$

$$E(y, \hat{y}) = \sum_{n=1}^{T} E(y_t, \hat{y}_t)$$

Now we have learn the data with the gradient $\frac{\partial E}{\partial y}$ back along the chain.
Also known as Back Propagation(stochastic descent method).
But the disadvantage is that the gradient vanishes along the chain.

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

LSTM - Long short Term Memory



These operations are used to allow the LSTM to keep or forget information.

It has forget gate which decides whether to hold the information or not.

This ensures that learning rate does not decrease down the chain.

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad