INTRO to AI And ML
Voice controlled Toy car

Ronak devda
EE19Btech11023
Dept. of Electrical Engg.,
IIT Hyderabad.

July 9, 2021

Zero Padding;-

This is a technique to increase no of samples of a given audio sample .
The necessity of padding is to generate more number of samples from given limited number of samples and each of uniform size. we pad them such that the recorded voice concentrates on the central part of padded samples without any noise added. we collected about 80 samples and made around 2000 samples of each speech.

Mel Frequency Cepstral coefficients

MFCC is a representation of the short-term power spectrum of a sound, which in simple terms represents the shape of the vocal tract. and when we load the data for training we compute mfcc's as vector of size [49,39]. and we label speech parts with some numbers.

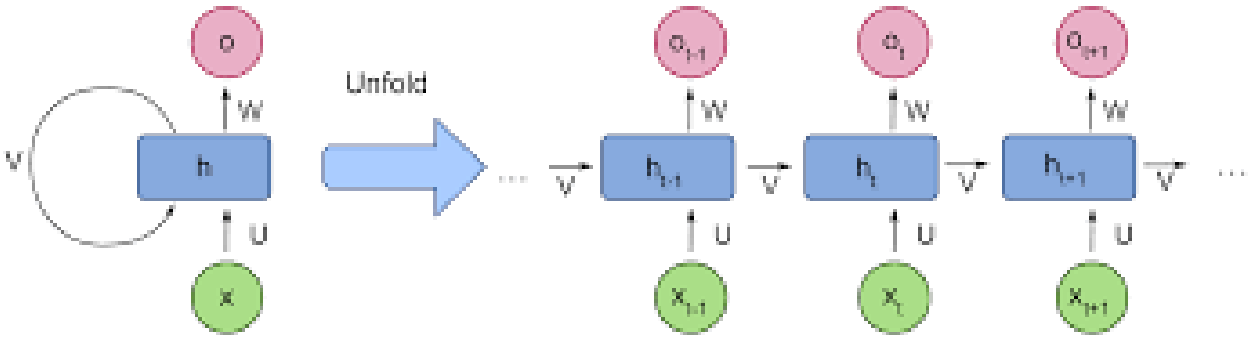
$$\text{Mel}(f) = 1125\ln(1 + f700)$$

CODE:-

```
back, sr = sf.read("Final/back"+str(i)+".wav")
back, index = librosa.effects.trim(back)
data.append(mfcc(y = back, sr = sr, nmfcc = 39).T)
label.append(0)
```

RNN (Recurrent Neural Network)

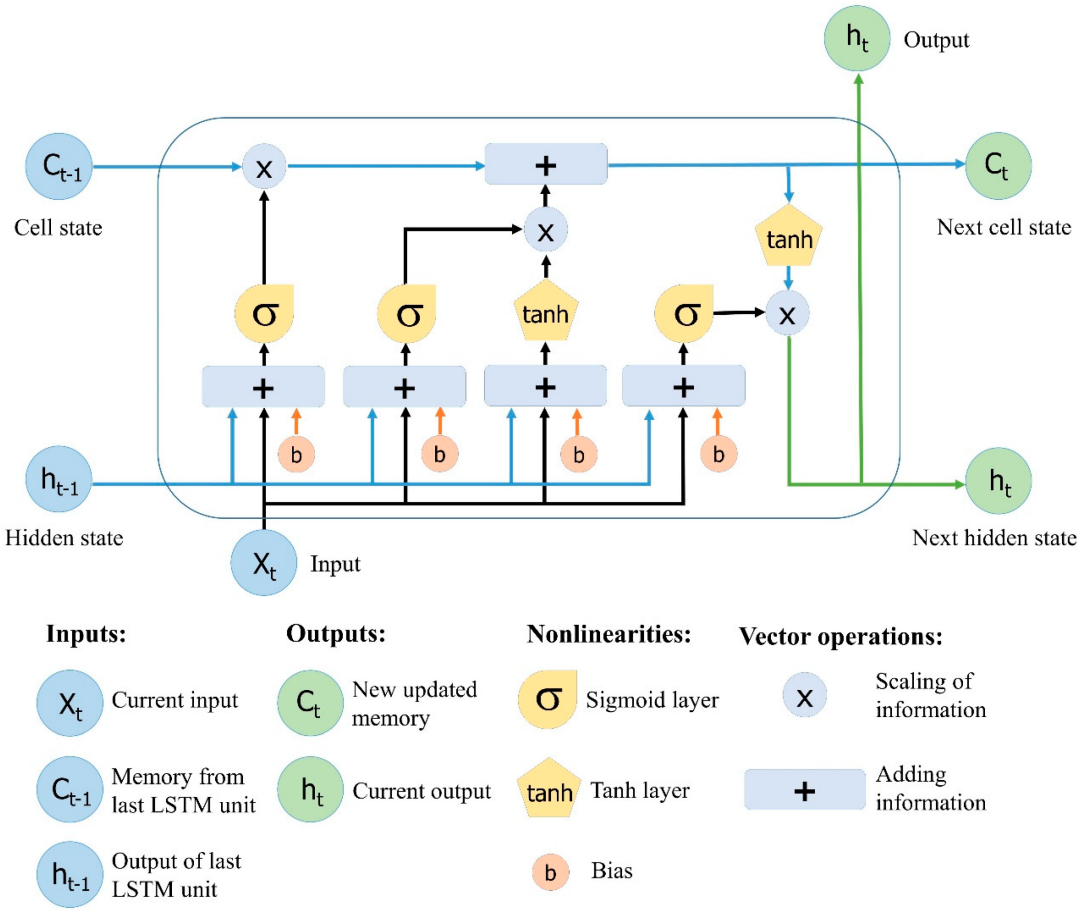
It is a generalization of feedforward neural network that has an internal memory.
These are type of Neural Networks in which output of previous state is taken as input to current state. The main and most important feature of RNN is Hidden state, which remembers some information about a sequence.



Long Short Term Memory

Standard RNNs suffer from vanishing and exploding gradient problems. LSTMs deal with these problems by introducing new gates, such as input and forget gates which enable better preservation of long-range dependencies.

What Actually LSTM is :-



Loss Function

The loss function in the LSTM is calculated as the categorical class entropy defined as follows

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

where C is the total number of classes and $\hat{y}_i$ is score of each sample of y_i in the softmax function.

Softmax function: $f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}}$