INTRODUCTION TO AI AND ML
Toycar
Presentation

P.Srijith Reddy,
EE19BTECH11041,
Dept. of Electrical Engg.,
IIT Hyderabad.

January 22, 2020

ZEROPADDING

MFCC's

RNN

Gradient Descent Method

P.Srijith Reddy, EE19BTECH11041, Dept. of Electrical Engg., IIT Hyderabad.

Zeropadding

Firstly we collect voice samples of required speech and then 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.

P.Srijith Reddy, EE19BTECH11041, Dept. of Electrical Engg., IIT Hyderabad.

Mel-frequency cepstral coefficients (MFCCs)

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.

$$Mel(f) = 1125 ln(1 + \frac{f}{700}) \tag{3.1}$$

CODE:

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


P.Srijith Reddy, EE19BTECH11041, Dept. of Electrical Engg., IIT Hyderabad.

RNN

Recurrent neural network A recurrent neural network (RNN) is a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence. RNNs can use their internal state (memory) to process sequences of inputs.

LSTM Long Short Term Memory networks usually just called LSTMs.

LSTM's are special kind of RNN's which can remember memory for long periods.

CODE

model.add(LSTM(units = 128, returnsequences = True, inputshape = (data.shape[1],39)))

P.Srijith Reddy, EE19BTECH11041, Dept. of Electrical Engg., IIT Hyderabad.

Gradient Descent Method

let's say x is input and y is original output and y' as output. Here for example forward is represented as

$$y = [1,0,0,0,0]^T and$$

similarly other parts. (5.1)

$$y' = sigmoid(W.X + B) \tag{5.2}$$

$$J(W,b) = 1/2(\sum (y - y')^2) \tag{5.3}$$

where

$$sigmoid(x) = (1/(1 + e^{-x})) \tag{5.4}$$

Here J(W,b) measures how close is output to original output and we need to minimize it's value.

P.Srijith Reddy, EE19BTECH11041, Dept. of Electrical Engg., IIT Hyderabad.

Loss function

J(W,b) can be minimized using gradient descent method.
The loss function here is let's say E

$$E = -\sum_{i=0}^{C} y_i \log(y_i') \tag{5.5}$$

(**CODE**

def categorical cross entropy(ytrue, ypred, axis=-1):
return -1.0*tf.reducemean(tf.reducesum(ytrue *
tf.log(ypred), axis))

P.Srijith Reddy, EE19BTECH11041, Dept. of Electrical Engg., IIT Hyderabad.

Softmax function

$$f(s_i) = \frac{e^{s_i}}{\sum_j^C e^{s_j}} \tag{5.6}$$

This function is used to generate outputs in the range of (0,1). eventually whose probabilities sum to 1

P.Srijith Reddy, EE19BTECH11041, Dept. of Electrical Engg., IIT Hyderabad.

P.Srijith Reddy,
EE19BTECH11041,
Dept. of Electrical Engg.,
IIT Hyderabad.