# Math In Voice Recognition

Mansi Nanavati
EE19BTECH11036

January 21, 2020

# Outline

# 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 $M$. Definition:

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

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. For example,

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

# Zero Padding

(i) Zero Padding helps in increasing the dataset collection by inserting zeros in the sequential data of soundfile.

(ii) The expanded dataset helps in creating many samples for the model to train.

(iii) This is done by 250files.py code.

# MFCC

(i) We imported mfcc from Librosa in Code.py to extract the features out of the soundfile.

(ii) We compute the mfcc as a feature vector of dimension (49,39).

(iii) 49 corresponds to the time steps and 39 the features in each time step.

# LSTM Neural Network

We use LSTM because we want to extract the features within the timesteps of soundfile and each phoneme is dependent on previous phoneme.
The model contains an LSTM layer to exploit the sequential nature of sound files.
Followed by maxpooling for eliminating the unnecessary information.
Then it is flattened and sent to the dense layer with 5 nodes with softmax as activation layer for probability prediction.

$x_t$ is the input at time step t.
$s_t$ is the hidden state at time step t. Its the memory of the network.
$s_t = f(U_{x_t} + W_{s_{t1}})$. The function $f$ usually is nonlinear such as tanh or ReLU. $o_1$ is the output at step t.

$$s_t = f(Ux_t + Ws_{t-1})$$
$$o_t = softmax(Vs_t)$$
$$E_t(y_t, \hat{y}_t) = -y_t log(\hat{y}_t)$$
$$E(y, \hat{y}) = \sum_t E_t(y_t, \hat{y}_t)$$

# Softmax

After reading data, it sends it in the format required. Labelling is done and classification task is done with the help of Softmax, estimating the probability at which it's doing it.
Softmax function, an activation function that turns numbers into probabilities that sum to one.
It outputs a vector that represents the probability distributions of a list of potential outcomes.

# Categorical Loss Function

Cross entropy loss, or log loss, measures the performance of the classification model whose output is a probability between 0 and 1. It increases as the predicted probability of a sample diverges from the actual value.
Cross entropy is defined as the following equation:

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

where C is the total number of classes.


Thank You