INTRODUCTION TO AI AND ML

B.Sana
EE19BTECH11021
Electrical Engneering

January 22, 2020

Outline

1. voice recognition
2. Generation of more samples
3. math involved
4. loss function
5. LSTM

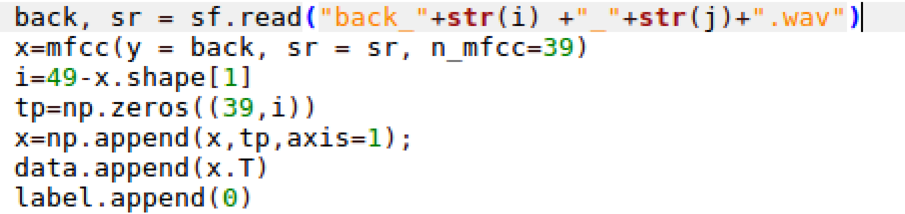
Generation of more samples

Create an empty array for new file to be generated from original file .
Then read the audio file which is already recorded using soundfile library .
By using a for loop add empty elements to the new_data array in start and at the end and the initial data in the middle. In this way an array of total length 25000 is created now the audio are written back to the hard disk.
Repeating this for loop generates more samples.

Math Involved

Create empty data and label array's. sf.read() reads the soundfile we record. then compute the MFCC as a feature vector of dimension (49,39) i.e. 49 time steps with each having 39 features.
Then the number of features it returns is 39 i.e. n_mfcc$=39$. it will of dimension(39,x)
as we don't know the number of zeros.
to find the number of zeros we will do (49-x).
y$=$ number of zeros to be added.

We will add the matrix x and y
and the transpose of the matrix will be stored in the empty data array
the data array is a 3D array with dimensions (9480,49,39).
the relation between the time steps can be exploited for the classification of the sound file to one of five classes.
https://github.com/sanabai/achieve/blob/master/25files.py



```
back, sr = sf.read("back_"+str(i) +"_"+str(j)+".wav")
x=mfcc(y = back, sr = sr, n_mfcc=39)
i=49-x.shape[1]
tp=np.zeros((39,i))
x=np.append(x,tp,axis=1);
data.append(x.T)
label.append(0)
```

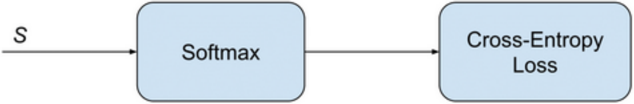
loss function

to_categorical converts a vector to binary vector with entries 0 and 1.
this is the implementation of categorical loss function from scratch
the input to the function is predicted probability and a one hot vector
it computes the loss by summing over the expression -ylog(p) over the tuple
this is summed over a batch and the final loss is given
i.e. categorical cross entropy:

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

where C is the total number of classes.

Categorical Cross-Entropy loss Also called Softmax Loss.
It is a Softmax activation plus a Cross-Entropy loss. If we use this loss, we will train a CNN to output a probability over the C classes for each image. It is used for multi-class classification.



$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}} \qquad CE = -\sum_i^C t_i log(f(s)_i)$$

LSTM

the model contains a LSTM layer to exploit the sequential nature of sound files Followed by maxpooling for eliminating the unnecesary information then it is flattened and sent to the dense layer with 5 nodes with softmax as activation layer for probability prediction