

Intro to AI and ML

Voice controlled bot

B Charitha Reddy
EE19BTECH11001
Electrical Engineering
IIT Hyderabad.

January 21, 2020

- 1 More Samples generation
- 2 Math in the code
- 3 softmax function
- 4 Categorical loss function
- 5 LSTM

More Samples generation

Create an empty array to generate new file from original file.

We have to then read the audio files recorded already using soundfile.

Use a for loop for generating more samples.

Add empty elements in the start and in the end to the array. Let the initial data elements be in the middle of the array.

So, an array of total length 25000 is created and audio files are written back to the hard disk.

Finally more voice files are created using the initial 80 voice files for each command.

Math in the code

Create data and label named empty arrays.

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

As there cannot be 49 time steps, fill the rest of the array (39,49-x) with 0's.

Then use np.append to add the two arrays with dimensions (39,x) and (39,49-x).

Now transpose it to get mfcc feature vector of (49,39) dimension.

https:

[//github.com/bojjacharitha/success/blob/master/25files.py](https://github.com/bojjacharitha/success/blob/master/25files.py)

Math in the code

After performing the MFCC on the input sound files we get an output matrix of 49x39 i.e. 49 time steps with each having 39 features.

The data is a 3-D array with (9480,49,39) .

label array is an array with the elements {0,1,2,3,4} as one row, similarly others are created.

```
back, sr = sf.read("back_"+str(i) + "_" + str(j) + ".wav")
```

```
x=mfcc(y = back, sr = sr, n_mfcc=39)
```

```
i=49-x.shape[1]
```

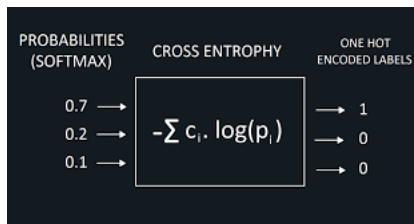
```
tp=np.zeros((39,i))
```

```
x=np.append(x,tp,axis=1);
```

```
data.append(x.T)
```

```
label.append(0)
```

softmax function



It squashes a vector in the range (0, 1) and all the resulting elements add up to 1.

$$f(\hat{y}_i) = \frac{e^{\hat{y}_i}}{\sum_j C e^{\hat{y}_j}}$$

Categorical loss function

`keras.utils.to_categorical` converts a class vector (integers) to binary class matrix.

It 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 . Categorical Cross entropy:

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

where C is the total number of classes.

LSTM

The model contains a lstm layer to exploit the sequential nature of sound files.

It is 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.

It is the layer for probability prediction.