**Gender Recognition using voice**

**Project Team:**

**Sarthak Kapse: 16D070041**

**Sumath Kandala: 16D100005**

**Hemant Kumawat: 16D100011**

**Mayur Gaikwad: 16D100008**

**1)Main Goal of project:-**

Recognition of gender using voice of the person.

**2)Approach:-**

We used a preprocessed dataset of 3168 different audio sample and finally we had a 20 different features for each examples and a label male or female. Initially I am converting male to 1 and female to 0 from our dataset . Then this features are feeded to a 6 layer neural network which has 20 units in input layer, [64,256,256,64] in respectively 4 hidden layer and a ouput layer consisting of 1 unit showing value between 0 to 1. And finally if value is greater than 0.5 we are converting it to male and if its less than or equal to 0.5 we are converting it to female. Dataset Link:- https://www.kaggle.com/primaryobjects/voicegender

**3)Experiments:-**

**a)Description**:- Our code for neural network is in python language using various libraries like pytorch, numpy, pandas, matplotlib and we used anaconda enviorment to run our code which is finally exported to py file for running in script with R file too. Our code for data interfacing i.e taking audio data and extracting its features and writing it in csv file is done In R using warbleR package and various other libraries. And finally this csv file is feed into our already trained neural network and is giving output male or female.

Our architecture of model we are using:-

Sequential(

(0): Linear(in\_features=20, out\_features=64, bias=True)

(1): ReLU()

(2): Dropout(p=0.25)

(3): Linear(in\_features=64, out\_features=256, bias=True)

(4): ReLU()

(5): Dropout(p=0.25)

(6): Linear(in\_features=256, out\_features=256, bias=True)

(7): ReLU()

(8): Dropout(p=0.25)

(9): Linear(in\_features=256, out\_features=64, bias=True)

(10): ReLU()

(11): Dropout(p=0.25)

(12): Linear(in\_features=64, out\_features=1, bias=True)

(13): Sigmoid()

)

And features of audio we are using are :-

meanfreq: mean frequency (in kHz)

sd: standard deviation of frequency

median: median frequency (in kHz)

Q25: first quantile (in kHz)

Q75: third quantile (in kHz)

IQR: interquantile range (in kHz)

skew: skewness (see note in specprop description)

kurt: kurtosis (see note in specprop description)

sp.ent: spectral entropy

sfm: spectral flatness

mode: mode frequency

centroid: frequency centroid (see specprop)

peakf: peak frequency (frequency with highest energy)

meanfun: average of fundamental frequency measured across acoustic signal

minfun: minimum fundamental frequency measured across acoustic signal

maxfun: maximum fundamental frequency measured across acoustic signal

meandom: average of dominant frequency measured across acoustic signal

mindom: minimum of dominant frequency measured across acoustic signal

maxdom: maximum of dominant frequency measured across acoustic signal

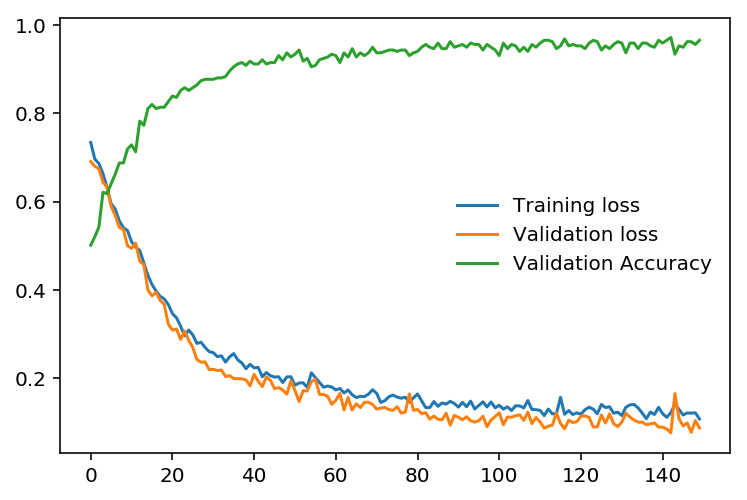
dfrange: range of dominant frequency measured across acoustic signal

modindx: modulation index. Calculated as the accumulated absolute difference between adjacent measurements of fundamental frequencies divided by the frequency range

label: male or female

**b) Expeimental Platform**:- We are running R code in Rstudio initially for getting features of a audio data into a csv file will be further used in neural network for recognising . Then we are running a file named training\_our\_model for training the model and finally saving all its trained weight and bias to checkpoint.pth. Then code file name testing\_our\_model is run in which we are initially loading weights and bais of learned model and loding model architecture too. Then the csv file created by Rstudio is passed through the network and we get our output which is either male or female.

**c)Experimental Results:**- From our dataset we are using 80% of data for training the model ,10% for validation and 10% for testing. We did hyperparameter tuning using validation loss and training loss plot (shown below). And we got epoch value of approx 150-200 and learning rate is 0.001 and droupout we are using is with p=0.25 and a ReLU after every layer and at output using sigmoid to get output in the range 0-1. While testing we saw that our accuracy is coming around 94-96% on test data that contain 10% of total examples from voice dataset we are using.



**4)Effort:-**

a)Most challenging part:- We were getting lot of problems while recording our input and running it through Rstudio and passing output csv to neural network to get the output.

b.) Work distribution:

Neural Nework : Sarthak Kapse, Hemant Kumawat and Sumanth Kandala

Audio Processing through R : Mayur Gaikwad and Sumanth Kandala

c.) Fraction of time spent on different work :

1. Literature reading: 10%
2. Neural network implementation and hyperparameters testing: 50%
3. Audio processing and through Rstudio : 40%

**5. Literature Read:**

Paper1: <https://www.researchgate.net/publication/312219824_Voice_Gender_Recognition_Using_Deep_Learning?fbclid=IwAR2o5ry6nc9AFof6TyKtgguhBjAZDg4RGs8NSxLTcwgN9RVHRL804XoC92U>

Paper 2 :

[http://cs229.stanford.edu/proj2014/Kunyu%20Chen,%20Gender%20Identification%20by%20Voice.pdf?fbclid=IwAR0Lwm7jn0Dq8XLjlsB53rh8Gda-rzuP3hLriWd2tPzMmYCzMcY7RGCqCPw](http://cs229.stanford.edu/proj2014/Kunyu Chen, Gender Identification by Voice.pdf?fbclid=IwAR0Lwm7jn0Dq8XLjlsB53rh8Gda-rzuP3hLriWd2tPzMmYCzMcY7RGCqCPw)