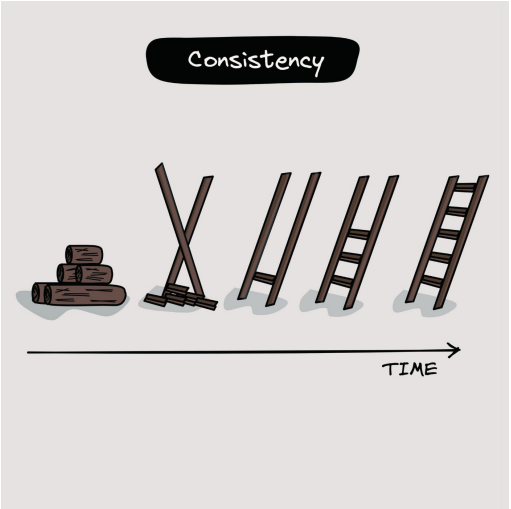
Case Study: Speech data and CNN

B.Tech. Data Science, NMIMS

By,

Bilal Hungund, Data Scientist, Halliburton



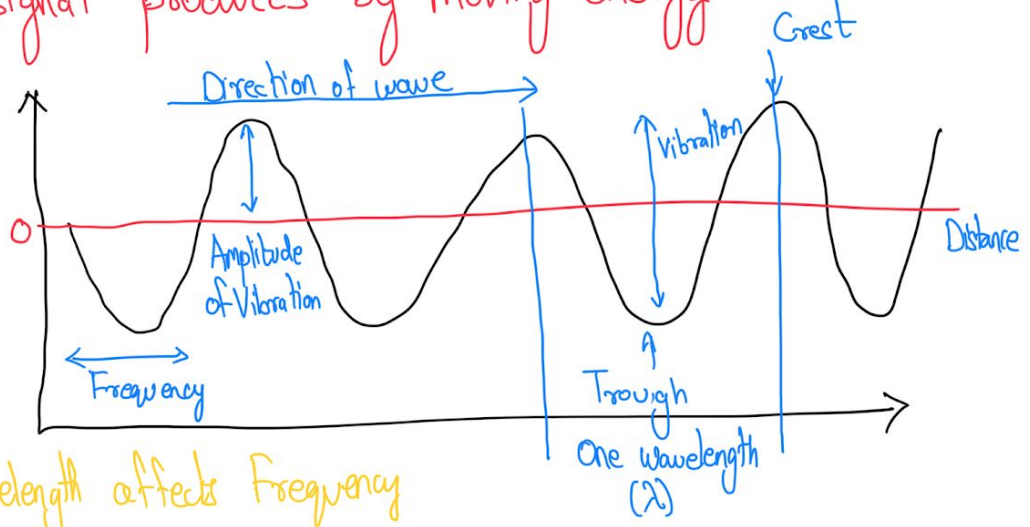
Audio Signal: (Automatic Speech Recognition)

Longitudinal vibration that produces vitality

Sound Wave:

Vibration signal produces by moving energy

Parameters
- Amplitude
- Crest and Trough
- Wavelength
- Cycle
- Frequency

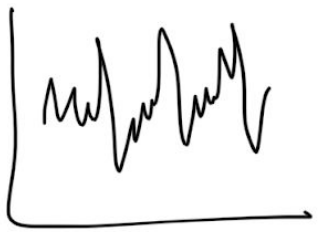


Wavelength affects Frequency

Acoustic Modelling

→ Statistical Representation of computed feature vectors

HMM — Segmental Model — Super Segment Model — Neural Network

 → Feature Extractor → Acoustic Model → Output

MFCC (Mel-frequency Ceptral Coefficients)

Mel Spectrogram

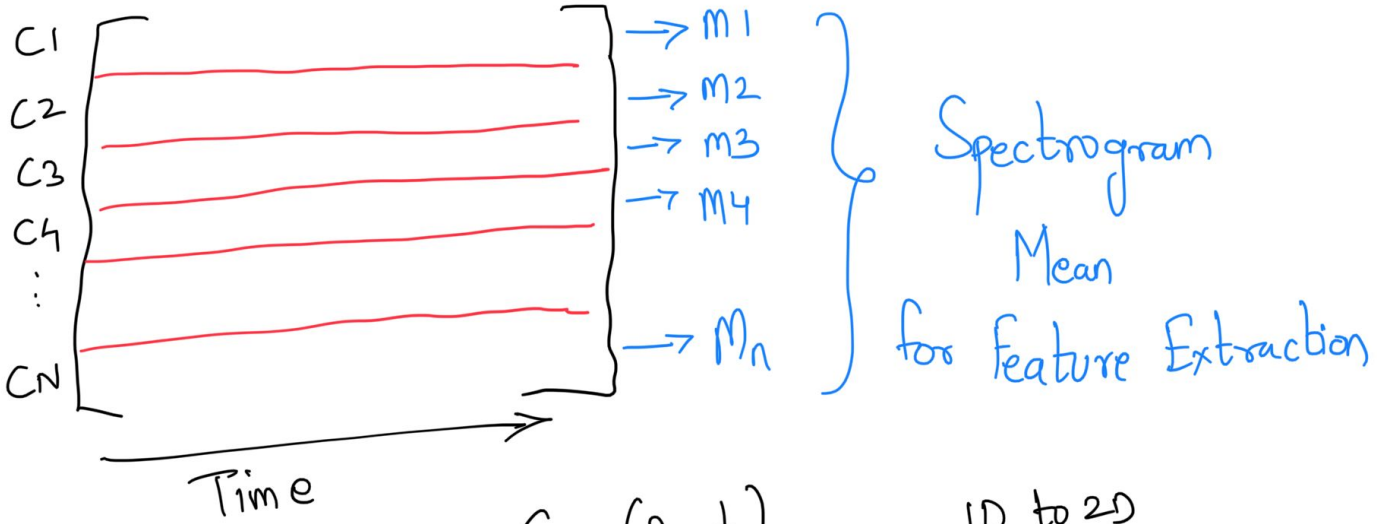
→ Spectrogram converted to Mel scale

→ Widely used in deep learning

→ Powerful tool to extract the feature from speech

→ Process includes: Fourier Transform, discrete cosine transforms and overlapping windows

→ It helps for classification problems such as genre classification, disease detection related to speech and etc.



$C = (n, t)$

1D to 2D

$M = (n, \ ) \longrightarrow (k, n_1, n_2, \#channels)$

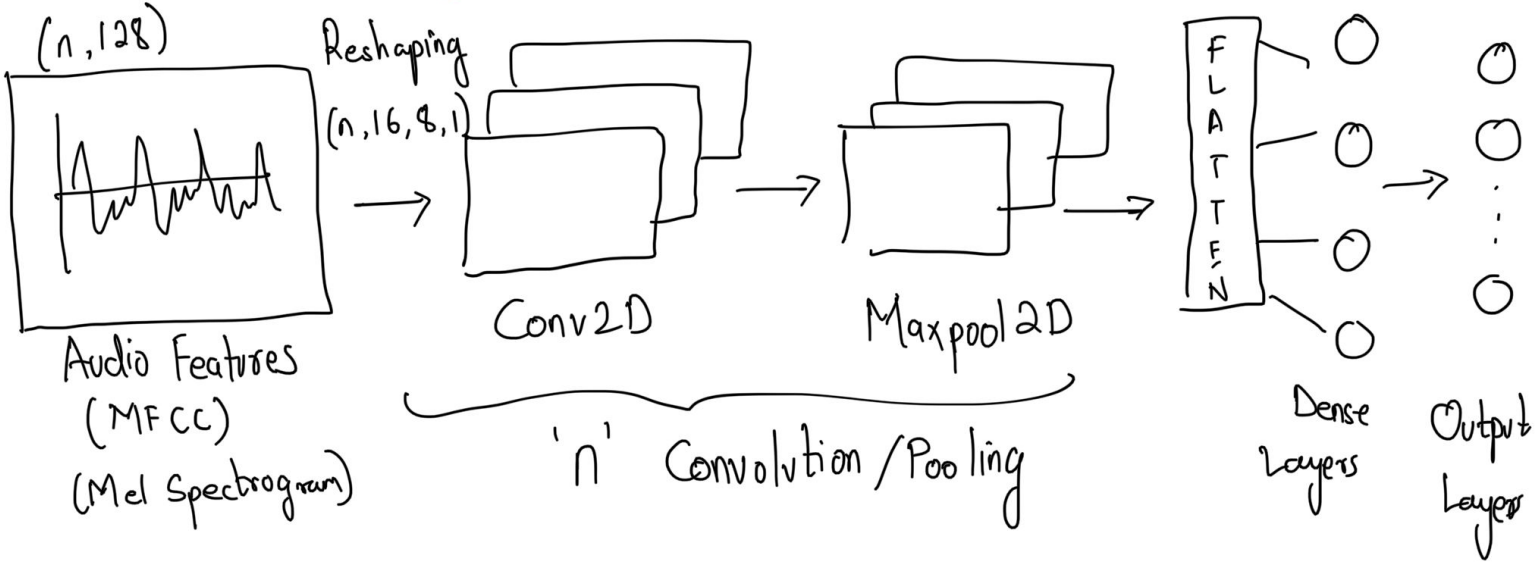
k → #samples

#channels ⇓ CNN Model

CNN in Speech Data

→ Create features using MFCCs & Mel Spectrogram

→ Average of matrix

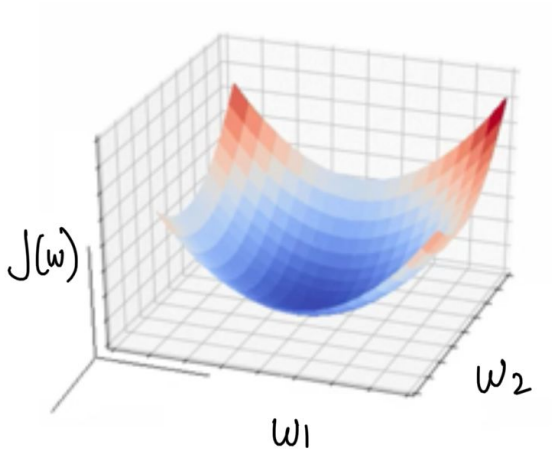


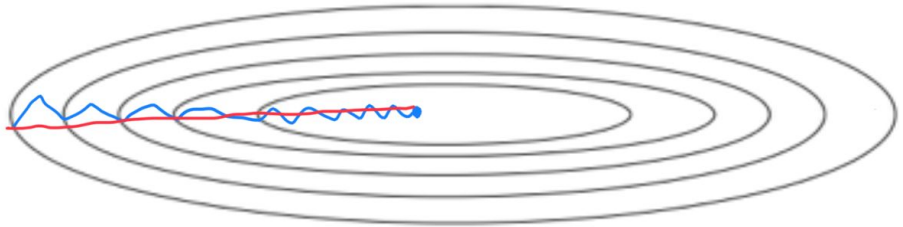
Adam Optimizer (Adaptive moment estimation)

- Momentum
- RMSprop

Exponentially weighted moving average

Momentum:





Momentum:

$$V_{dW} = \beta V_{dW} + (1-\beta)\, dW$$

$$V_{dB} = \beta V_{dB} + (1-\beta)\, dB$$

$$W := W - \alpha V_{dW}$$

$$B := B - \alpha V_{dB}$$

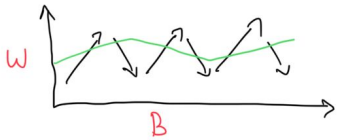
$[\alpha$ = learning rate$]$

$$W := W - \alpha\, dW$$

$$B := B - \alpha\, dB$$

$\beta = 0.9$

RMS prop (Root mean squared proportion)



$$S_{dW} = \beta S_{dW} + (1-\beta)(dW)^2$$

$$S_{dB} = \beta S_{dB} + (1-\beta)(dB)^2$$

$$W := W - \alpha \frac{dW}{\sqrt{S_{dW} + \varepsilon}}$$

$$B := B - \alpha \frac{dB}{\sqrt{S_{dB} + \varepsilon}}$$

$\beta = 0.999$

$\varepsilon = 10^{-8}$

Adam:

$$W := W - \alpha \frac{V_{dW}}{\sqrt{S_{dW} + \varepsilon}} \qquad B := B - \alpha \frac{V_{dB}}{\sqrt{S_{dB} - \varepsilon}}$$

Momentum β will be β_1 | RMS prop β will be β_2