Thursday Learning Hour

Music and Deep learning

Prabakaran Chandran

Agenda:

- Music Generation
- Music Classification

How does Machine see Music?

Music can be notated in multiple ways , Today we will discuss two among them (out of Many ways we have selected these 2 to perform the hands-on tasks)

1. ABC Notation – We will use this for Music Generation – Textual Representation
2. Spectrogram – We will use this for Music Genre Classification – Visual Representation

What is ABC Notation?

Music Directors' Way of Notation

Other than this , Computer readable Versions are available

1.MIDI – Music Instrument Device Instrument
2.ABC Notation

The Above one is Sheet Representation . Here ,music is represented by a sequence of musical notes. Each musical note is separated by a space. This can be used to represent both single instrument and multi-instrument music.

How does Machine see Music? - ABC Notation

```
<score lang="ABC">
X:1
T:The Legacy Jig
M:6/8
L:1/8
R:jig
K:G
GFG BAB | gfg gab | GFG BAB | d2A AFD |
GFG BAB | gfg gab | age edB |1 dBA AFD :|2 dBA ABd |:
efe edB | dBA ABd | efe edB | gdB ABd |
efe edB | d2d def | gfe edB |1 dBA ABd :|2 dBA AFD |]
</score>
```

Part-1

Part-2

1. Convert Each Characters (Notes) in to One hot encoded Embedding
2. So that our Machine will understand the Notes and Process them
3. EX: G → 00000000100000000
4. There are 95 unique characters in the dataset – (My Kaggle Kernel)

ABC notation is a shorthand form of musical notation for computers. In basic form it uses the letter notation with a–g, A–G, and z, to represent the corresponding notes and rests, with other elements used to place added value on these – sharp, flat, raised or lowered octave, the note length, key, and ornamentation

Part-1 represents meta data. Lines in the Part-1 of the tune notation, beginning with a letter followed by a colon, indicate various aspects of the tune such as the index, when there are more than one tune in a file (X:), the title (T:), the time signature (M:), the default note length (L:), the type of tune (R:) and the key (K:).
Part-2 represents the tune, which is a sequence of characters where each character represents some musical note.

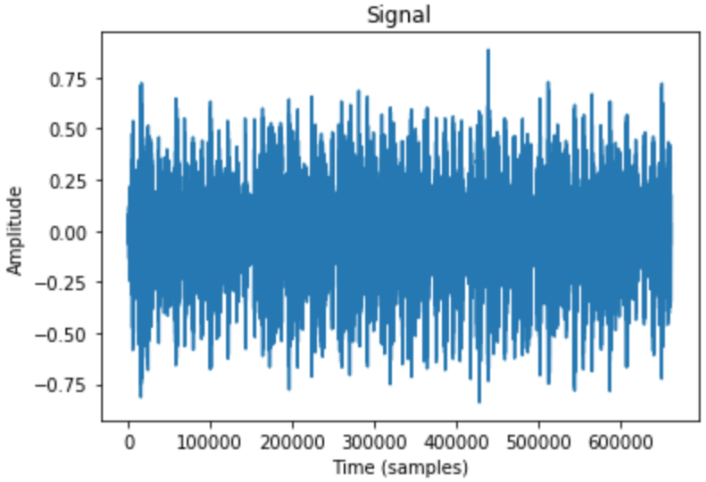
How does Machine see Music? -- Spectrogram

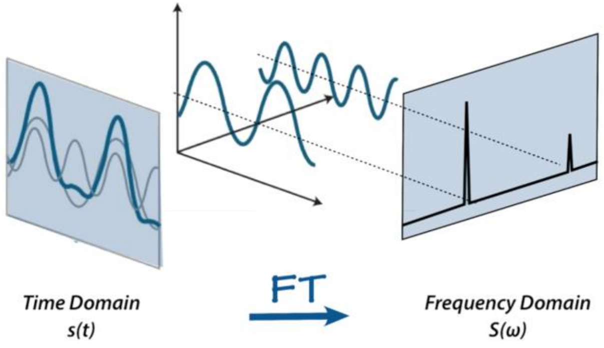
1. A **signal** is a variation in a certain quantity over time. For audio, the quantity that varies is air pressure. How do we capture this information digitally?

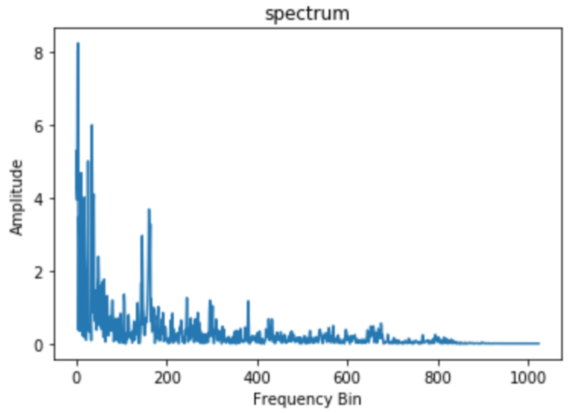
2. We can take samples of the air pressure over time. The rate at which we sample the data can vary, but is most commonly 44.1kHz, or 44,100 samples per second. What we have captured is a **waveform** for the signal, and this can be interpreted, modified, and analysed with computer software.

3. To Extract More information , This should not be in time domain ,FT will help us to decompose the Time domain Wave form

4. The **Fourier transform** is a mathematical formula that allows us to decompose a signal into it's individual frequencies and the frequency's amplitude. In other words, it converts the signal from the time domain into the frequency domain. The result is called a **spectrum**.







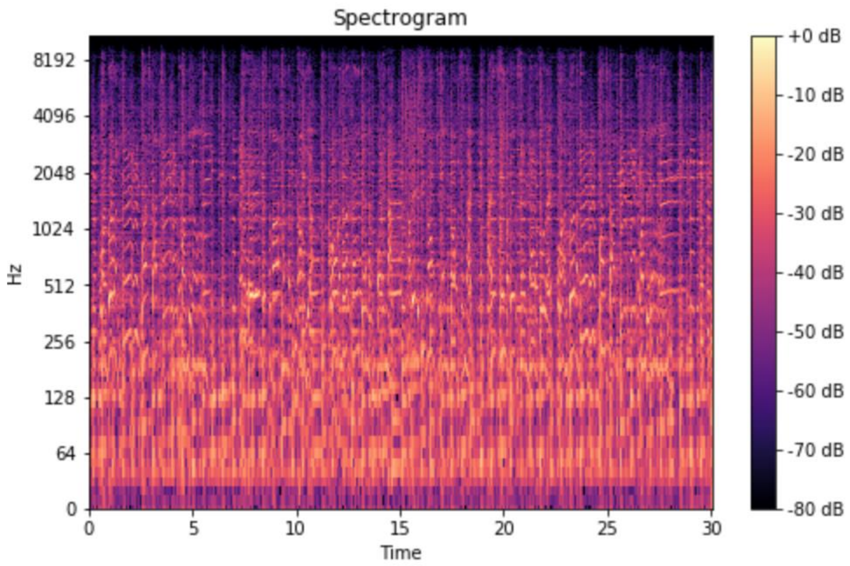
How does Machine see Music? -- Spectrogram

The fast Fourier transform is a powerful tool that allows us to analyse the frequency content of a signal, but what if our signal's frequency content varies over time?

Such is the case with most audio signals such as music and speech. These signals are known as **non periodic** signals.

We need a way to represent the spectrum of these signals as they vary over time. You may be thinking, "hey, can't we compute several spectrums by performing FFT on several windowed segments of the signal?"

Yes! This is exactly what is done, and it is called the **short-time Fourier transform.** The FFT is computed on overlapping windowed segments of the signal, and we get what is called the **spectrogram.**



Deep learning and Music Generation

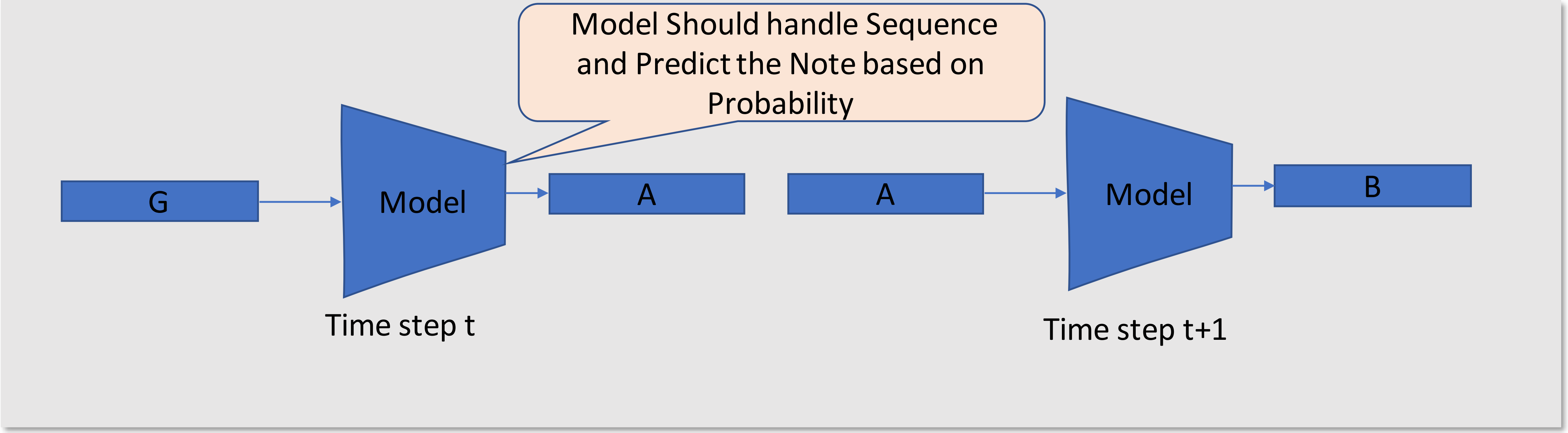
Task : To Generate the Musical Notes

Data set : 100s of ABC Notations

Data type : Sequence

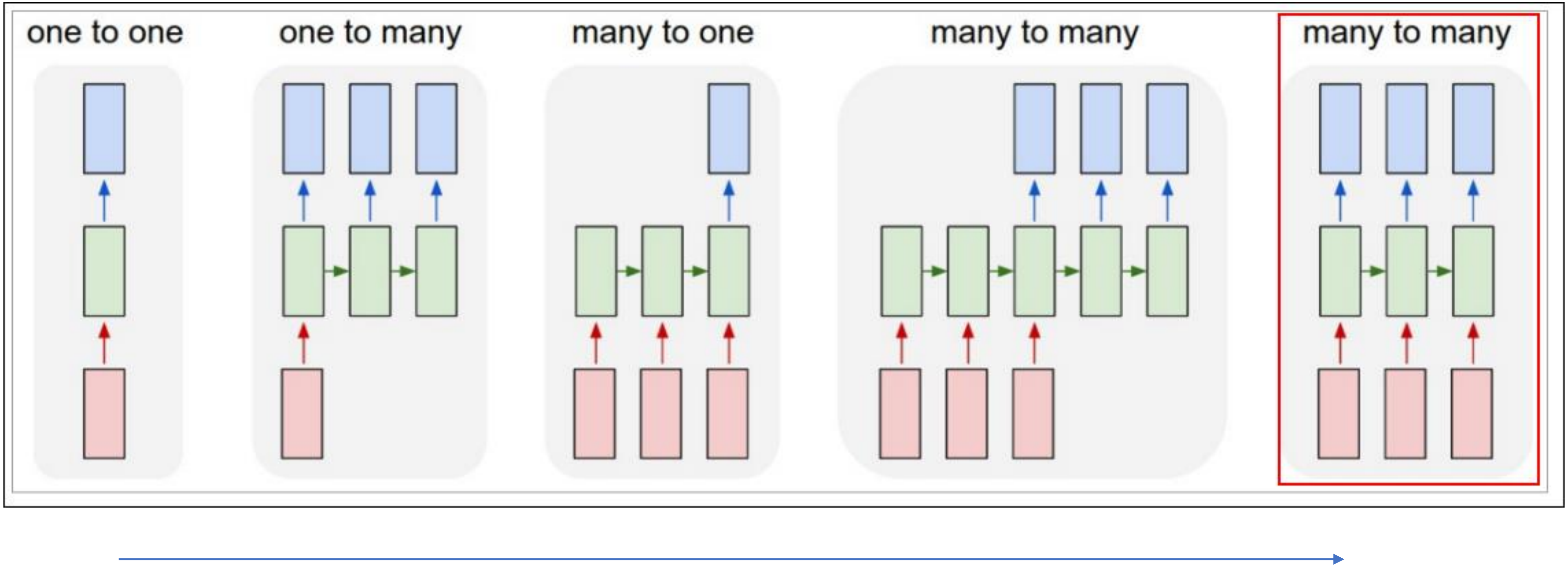
Preferred Neural Architecture : From the Boltzmann Family – Recurrent Neural Network

Overview:



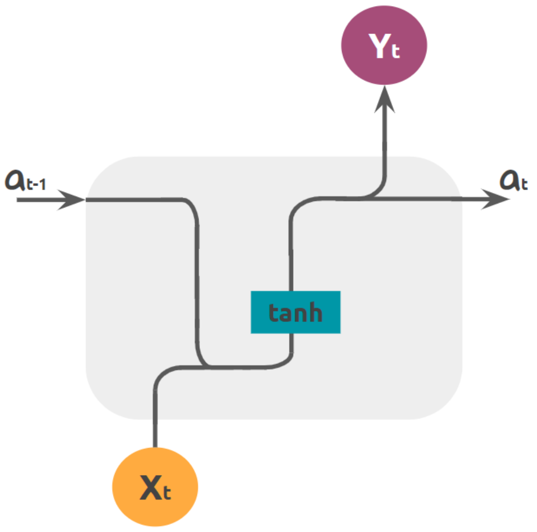
Model Architecture to Process the Music Sequence

- Recurrent Neural Networks are the Fundamental Architecture which are used for Sequential tasks (1980s)
- Many Variations in Recurrent Neural Networks are available , For this one We will use Many to Many RNN Architecture.

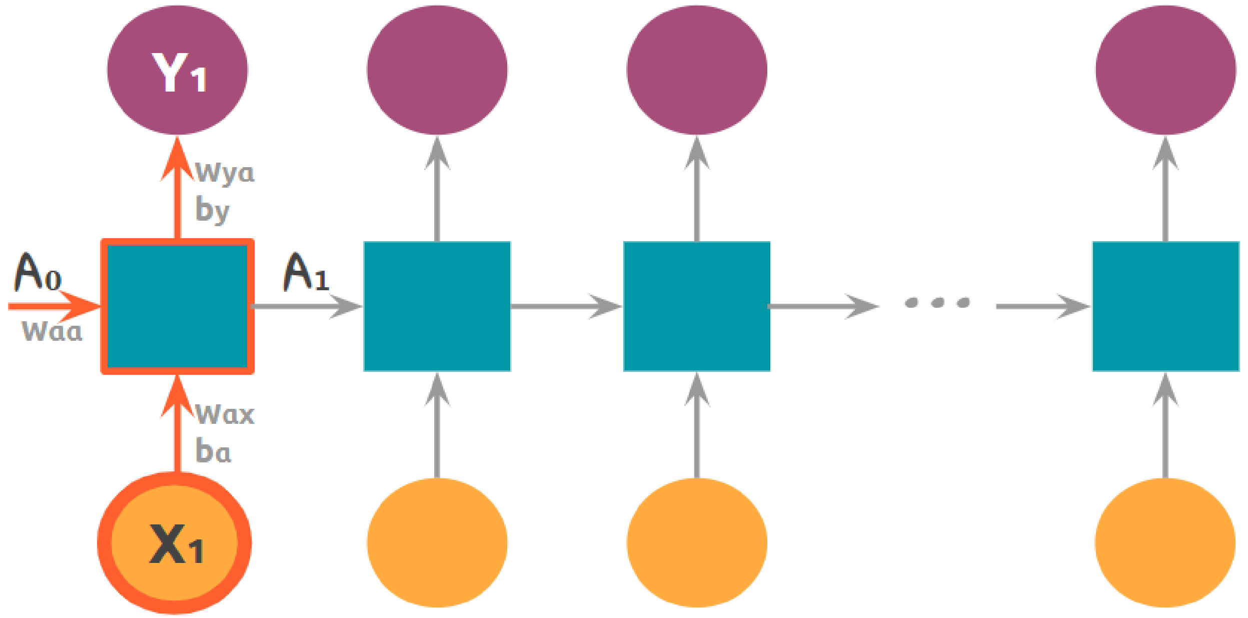


Time Variant Architectures

Deep learning and Music Generation – Char RNN Architecture



‹ RNN ›



< RNN >

$A_t = \tanh(W_a \cdot [A_{t-1}, X_t] + b_a)$

$Y_t = g(W_y \cdot A_t + b_y)$

- When it was ANN,

$Z_1 = W_1 \cdot X + b_1$

$A_1 = g(Z_1)$

→

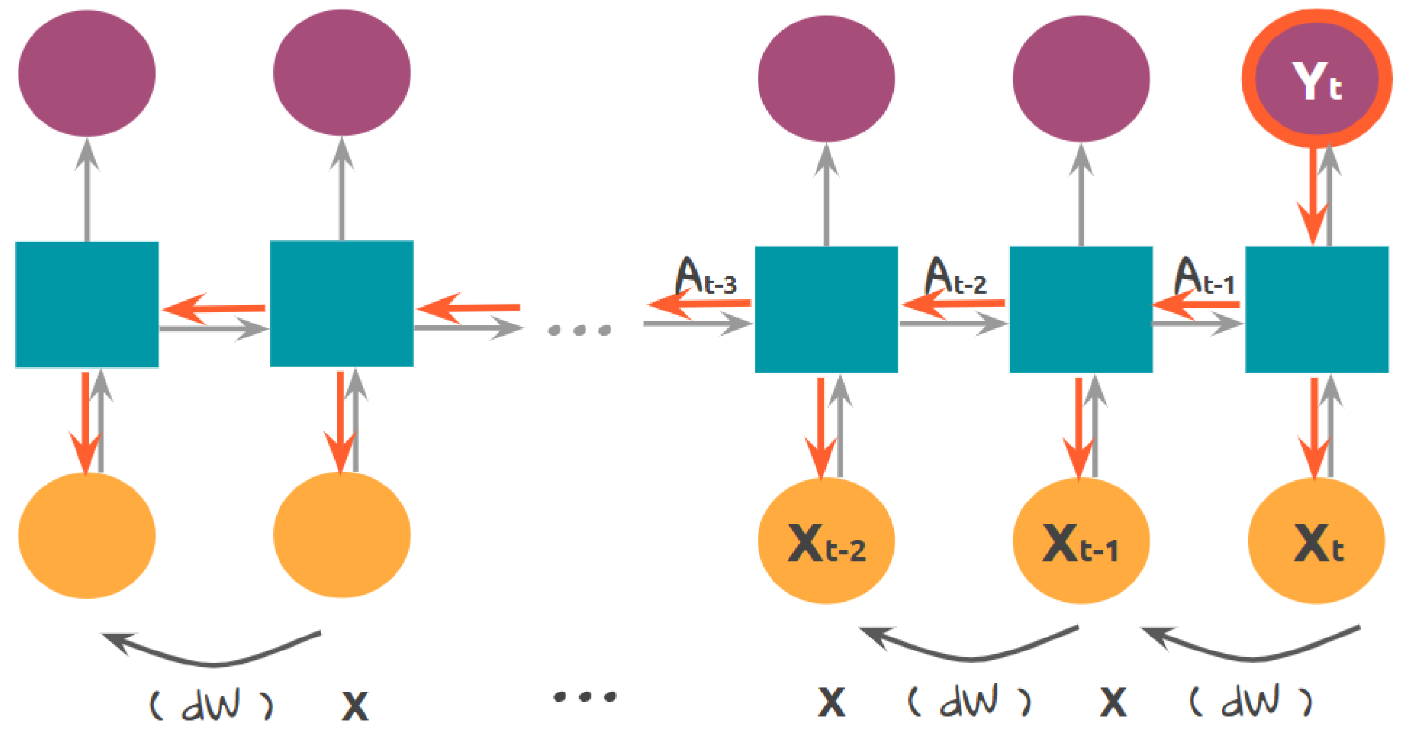
- Now in RNN,

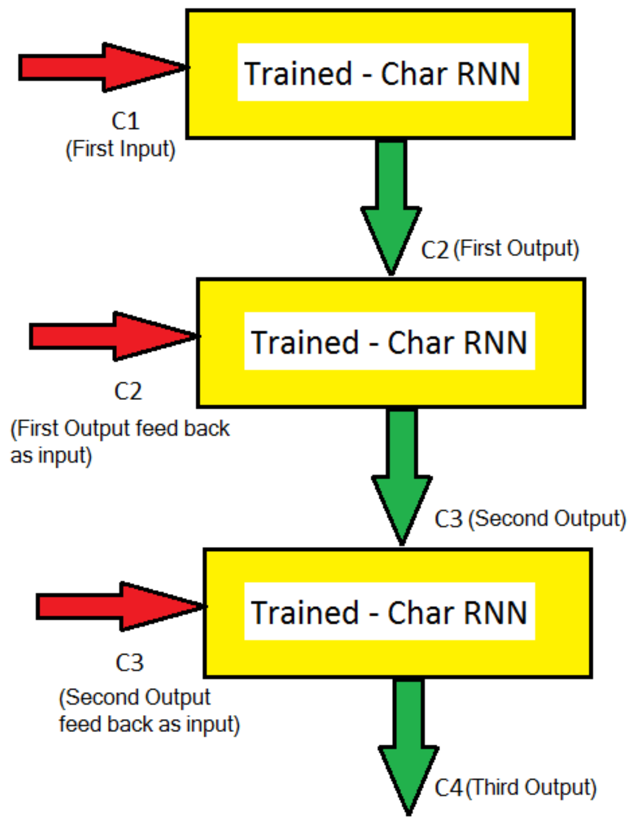
$Z_1 = W_{ax} \cdot X_1 + W_{aa} \cdot A_0 + b_a$ ···· ①

$A_1 = g(Z_1)$ ···· ②

$Y_1 = g(W_{ya} \cdot A_1 + b_y)$ ···· ③

Deep learning and Music Generation – Char RNN Architecture





Deep learning and Music Genre Classification

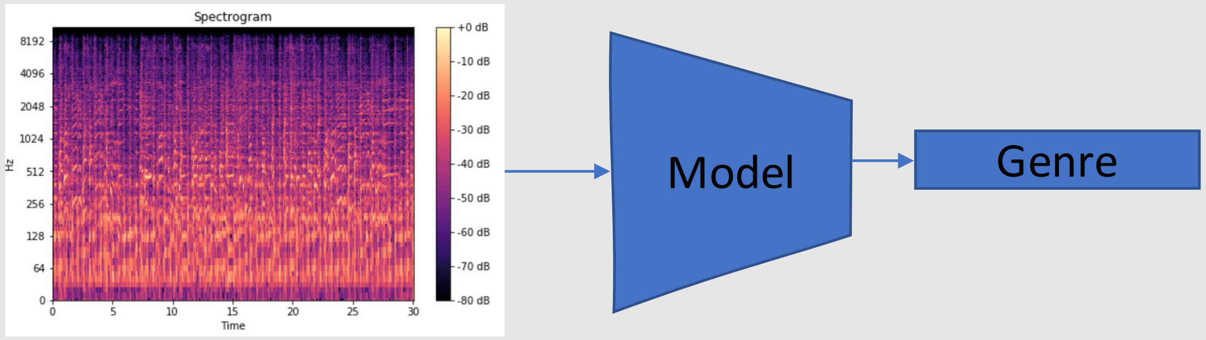
Task : To Classify the Music Genre Using Wave files

Data set : 1000s of wave file – Spectrogram(Conversion)

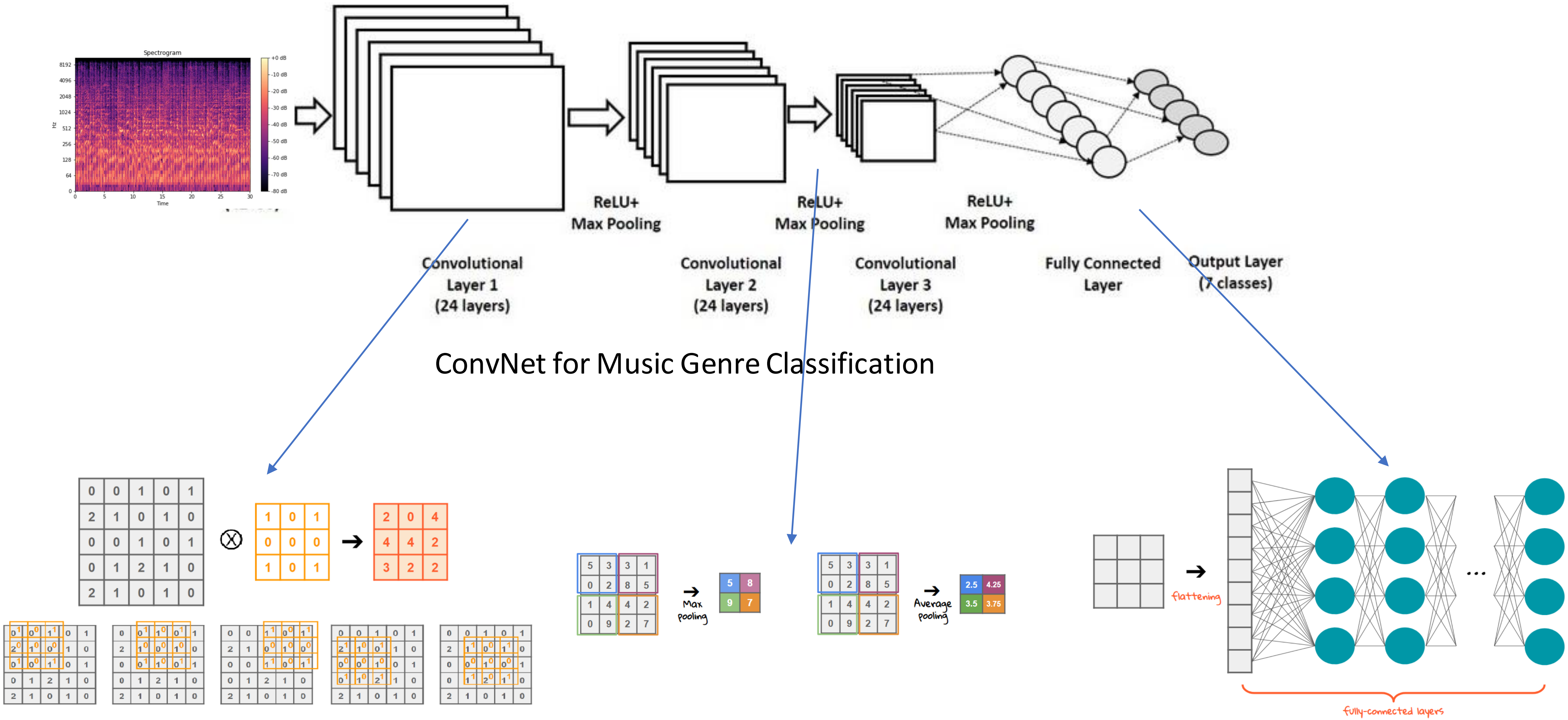
Data type : Visual Representation - Image

Preferred Neural Architecture : Convolutional Neural Network – either Custom network or Image Net Architecture.

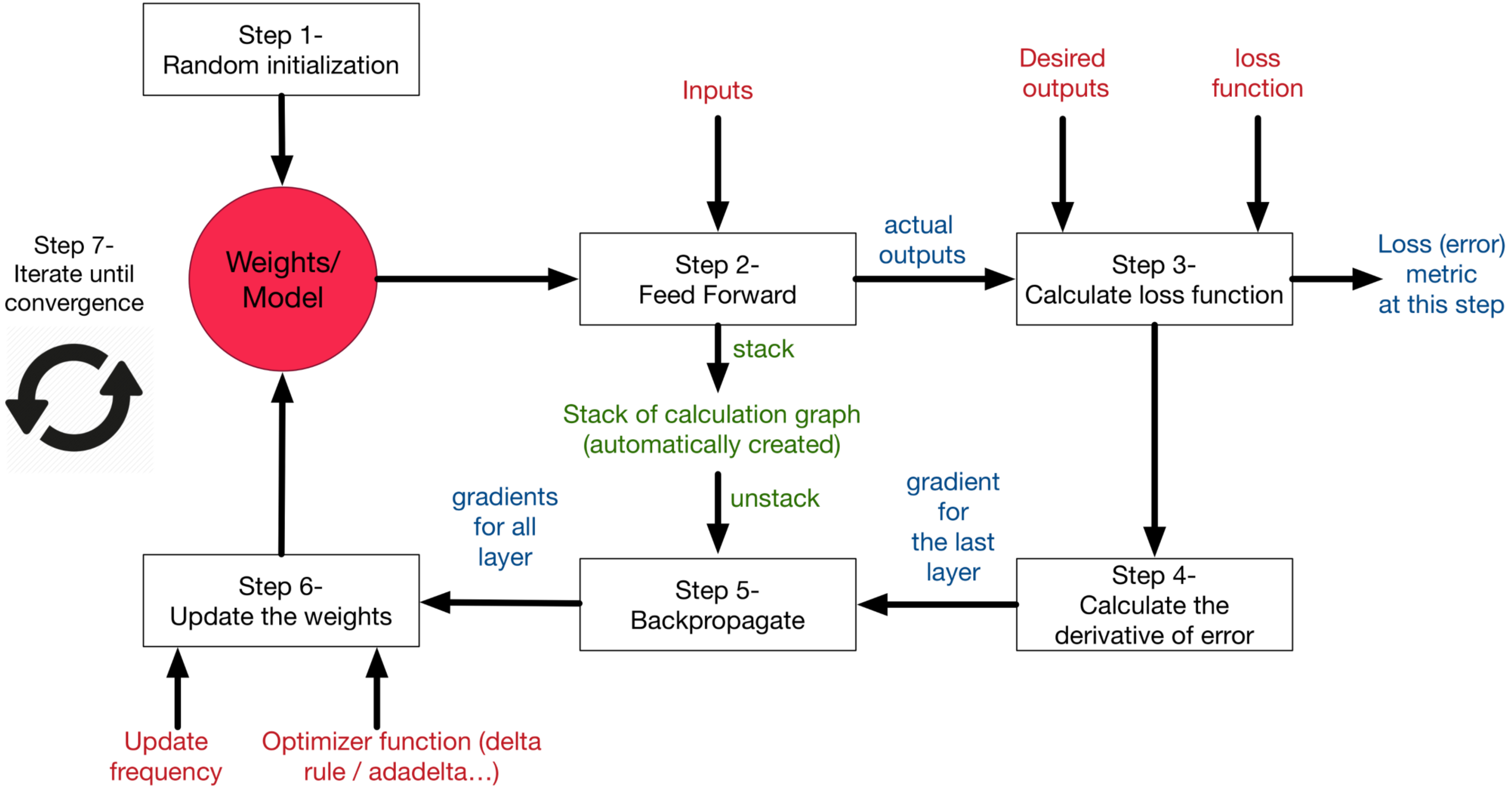
Overview:



Deep learning and Music Genre Classification – ConvNet

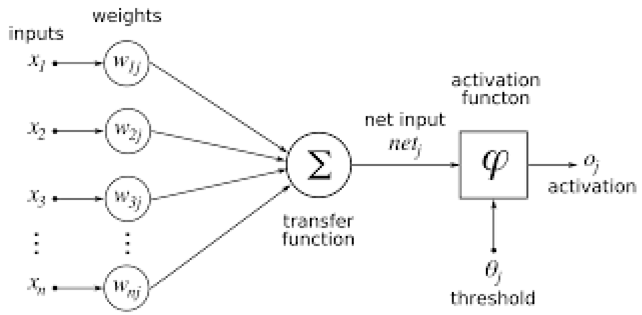


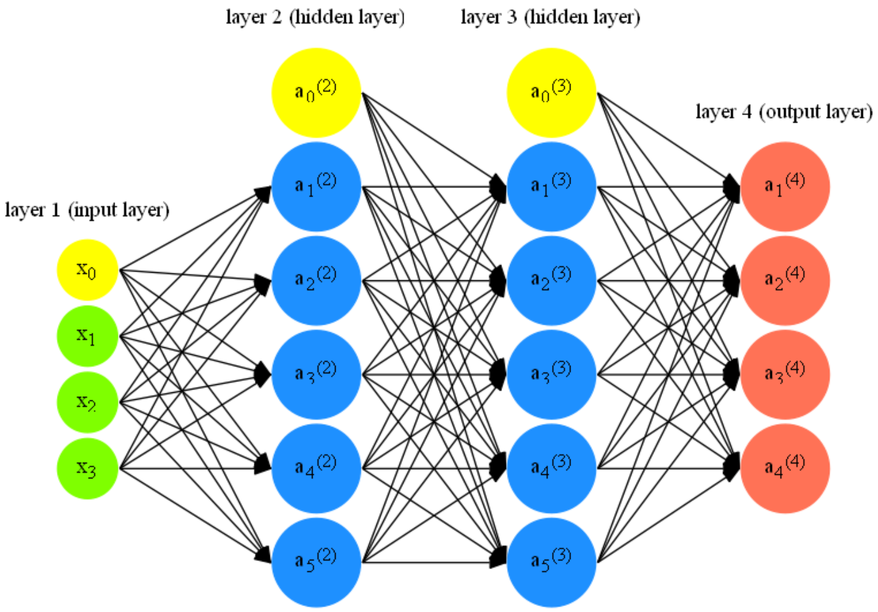
How does a DNN learn things? – Work flow for any Tasks

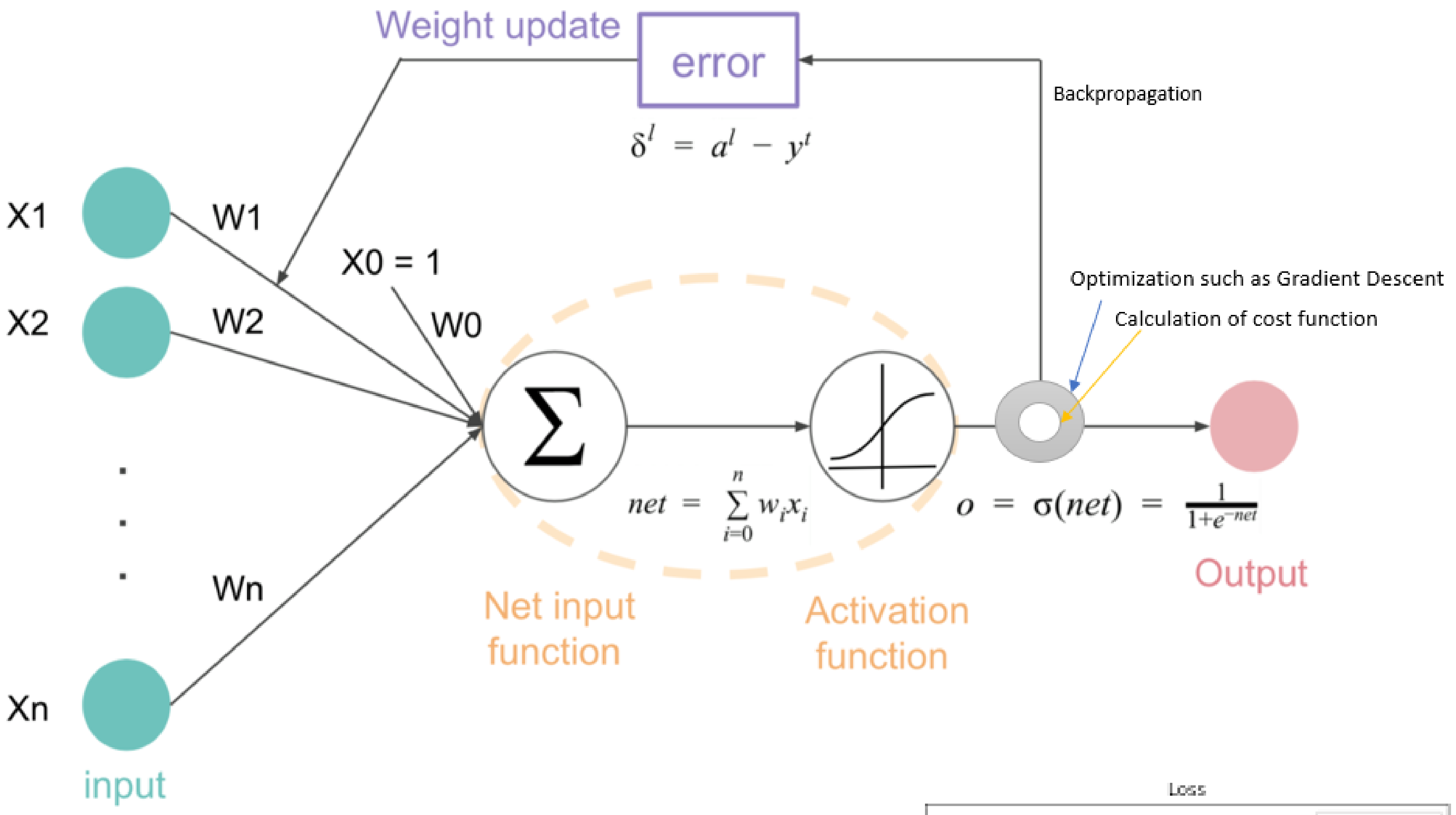


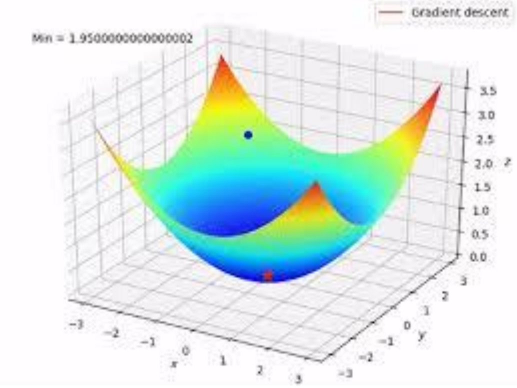
Appendix

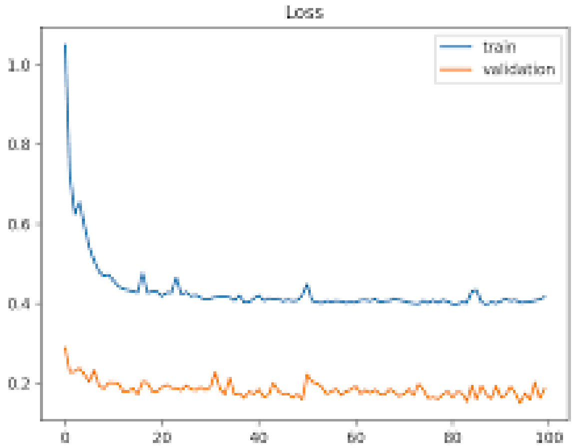
Deep Learning : overview



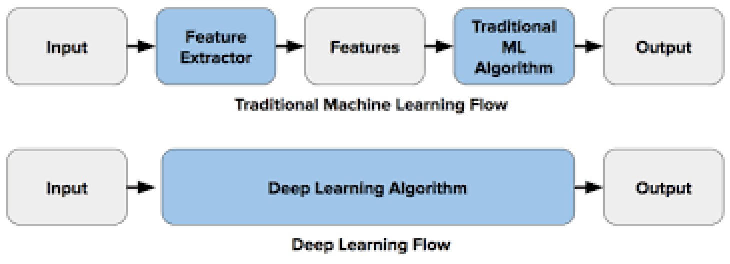




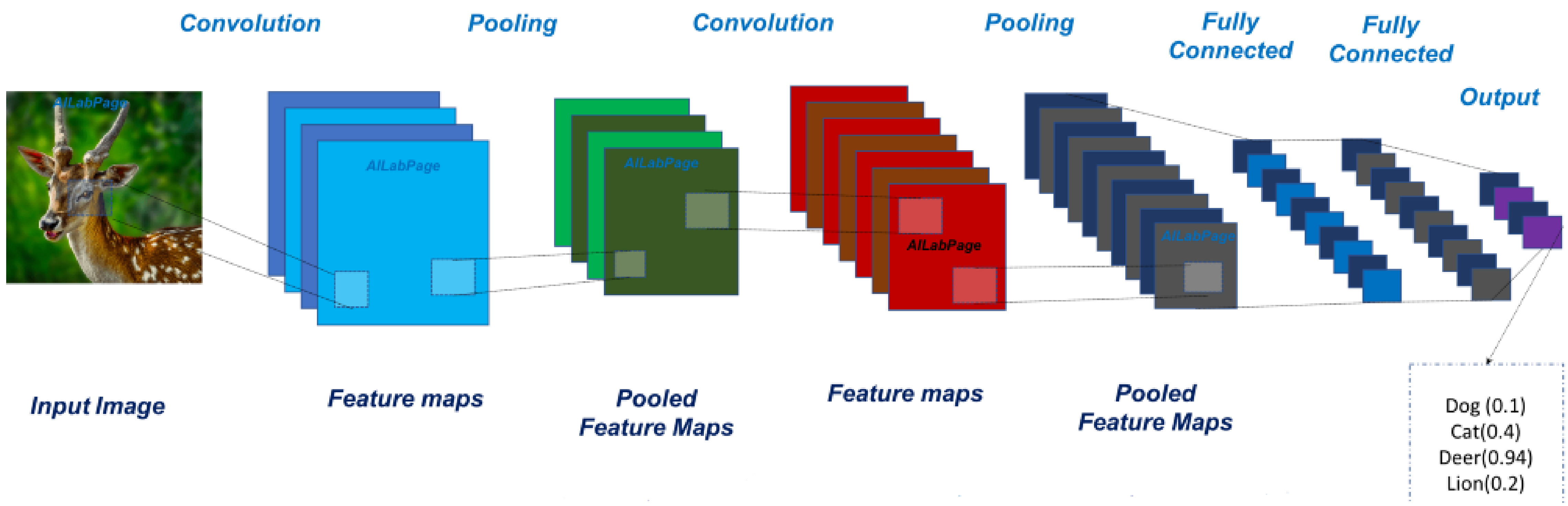




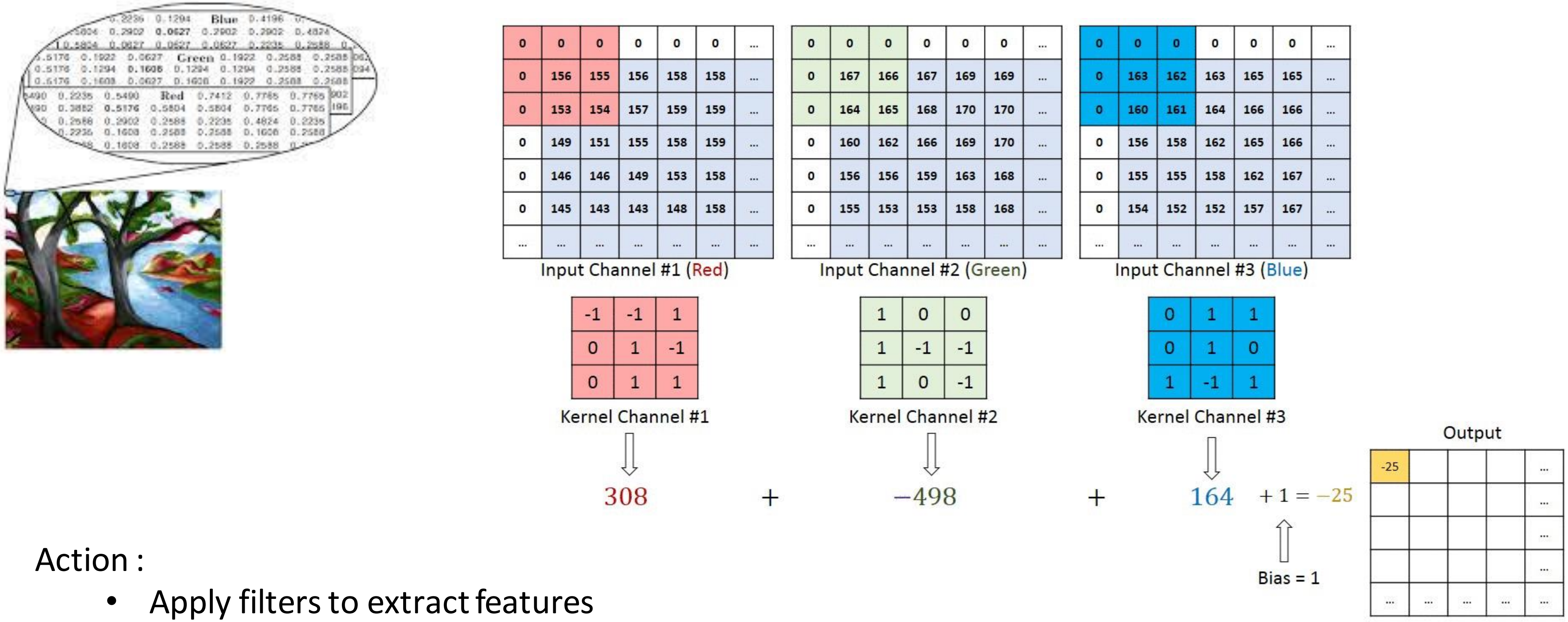
Computer Vision + Deep learning : Convolutional Neural Network.



Traditional way of CV extracts features using some explicit tasks like wavelet transformation , image processing , but in Deep learning everything is being handled by the network itself.



Convolutional Neural Networks – different layers of CNN



Action :

- Apply filters to extract features
- Filters are composed of small kernels, learned
- One bias per filter
- Apply activation function on every value of feature map

Parameters

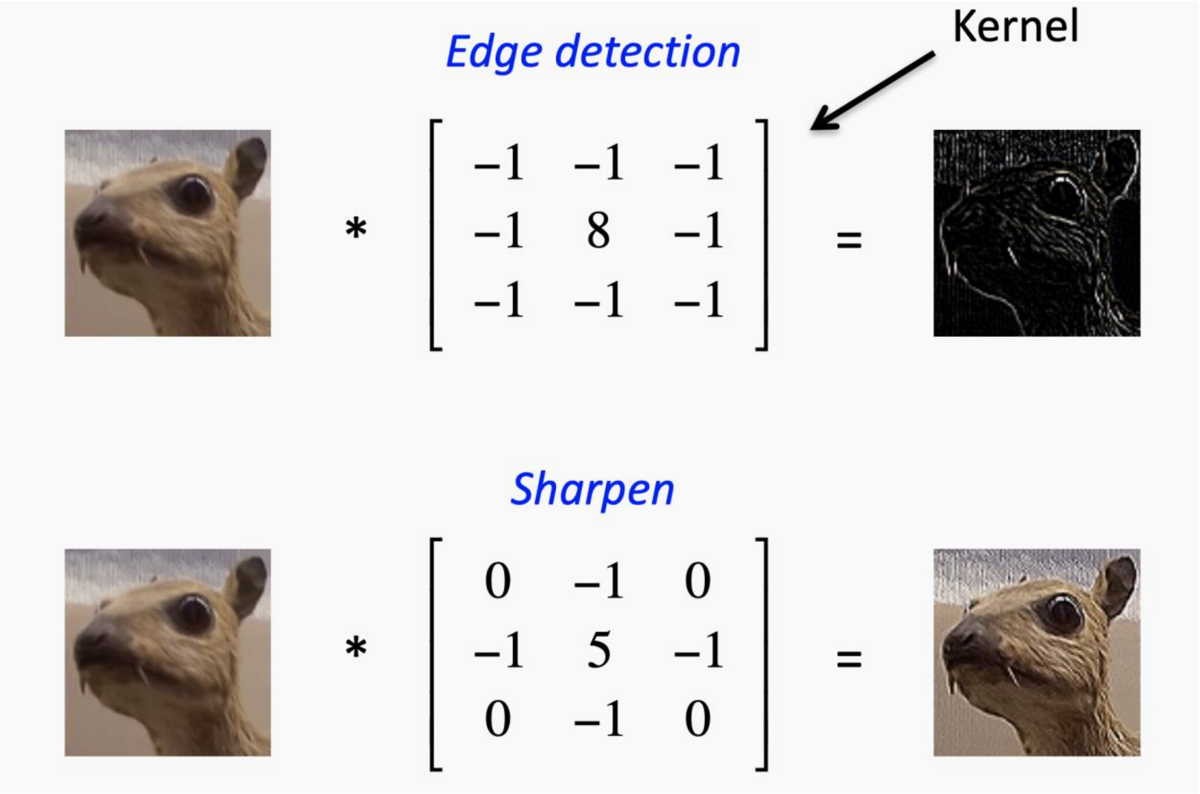
- Number of Kernels , size of kernels
- Activation function ,striding , padding

Convolutional Neural Networks – different layers of CNN

Before and after Convolution

52	34	14	5
45	12	17	11
29	20	19	27
99	85	60	55
120	112	88	29

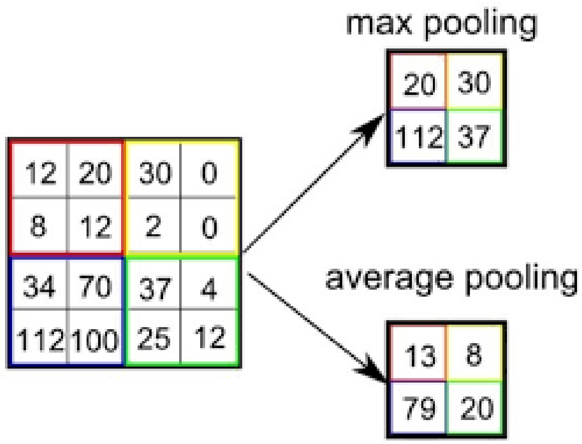
131	122	97	17
114	104	78	88
74	144	99	112
156	213	113	98
146	177	120	130



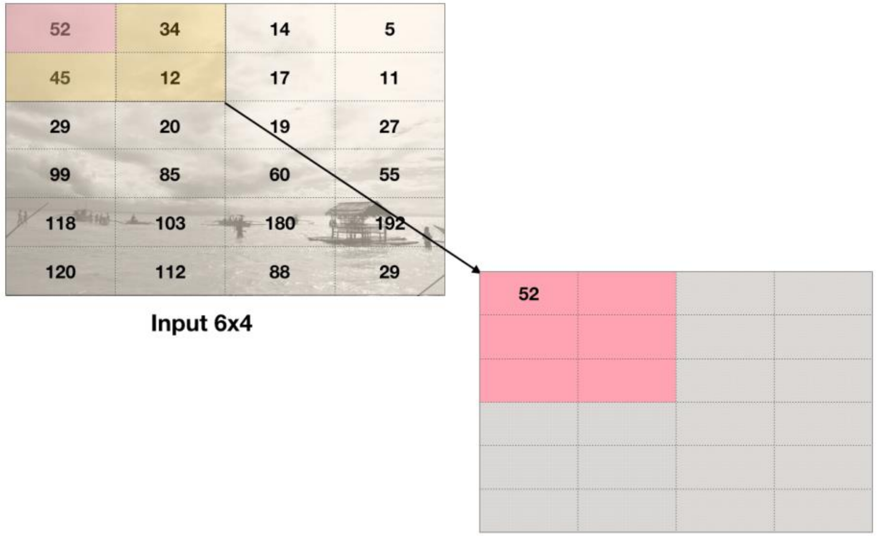
$$* \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} =$$

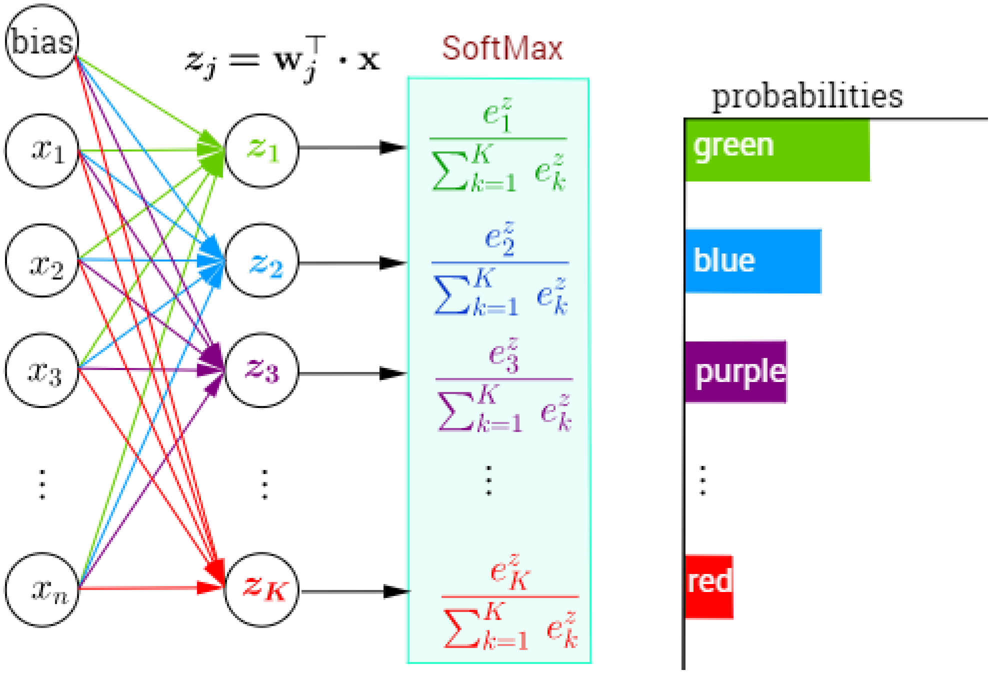
$$* \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} =$$

Convolutional Neural Networks – different layers of CNN



Max Pooling





Pooling layers:

- Reduce Dimensionality
- Extract maximum of / average of a region
- Follow Sliding window approach

Fully connected layers:

- Aggregate information from final feature maps
- Flatten the feature maps for final classification
- Generate final classification with the use of Sigmoid/SoftMax

https://medium.com/artists-and-machine-intelligence/neural-nets-for-generating-music-f46dffac21c0

medium.com/artists-and-machine-intelligence/neural-nets-for-generating-music-f46dffac21c0

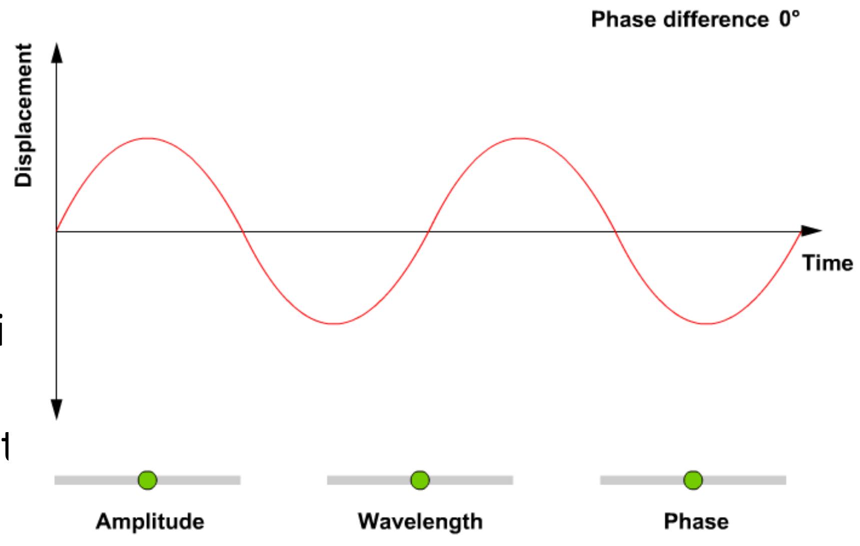
https://medium.com/@mheavers/machine-learning-in-sound-music-6f0715320d49

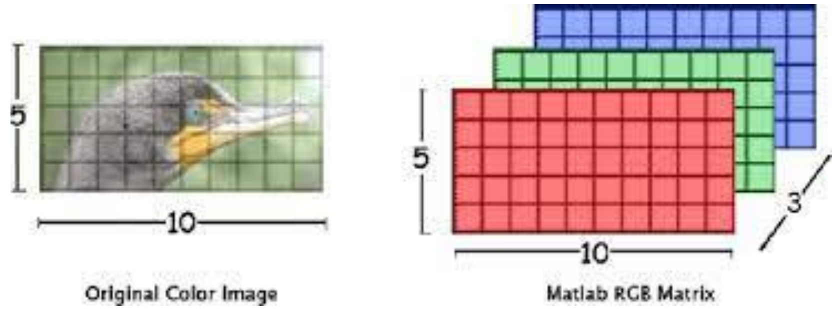
How does Machine see Music?

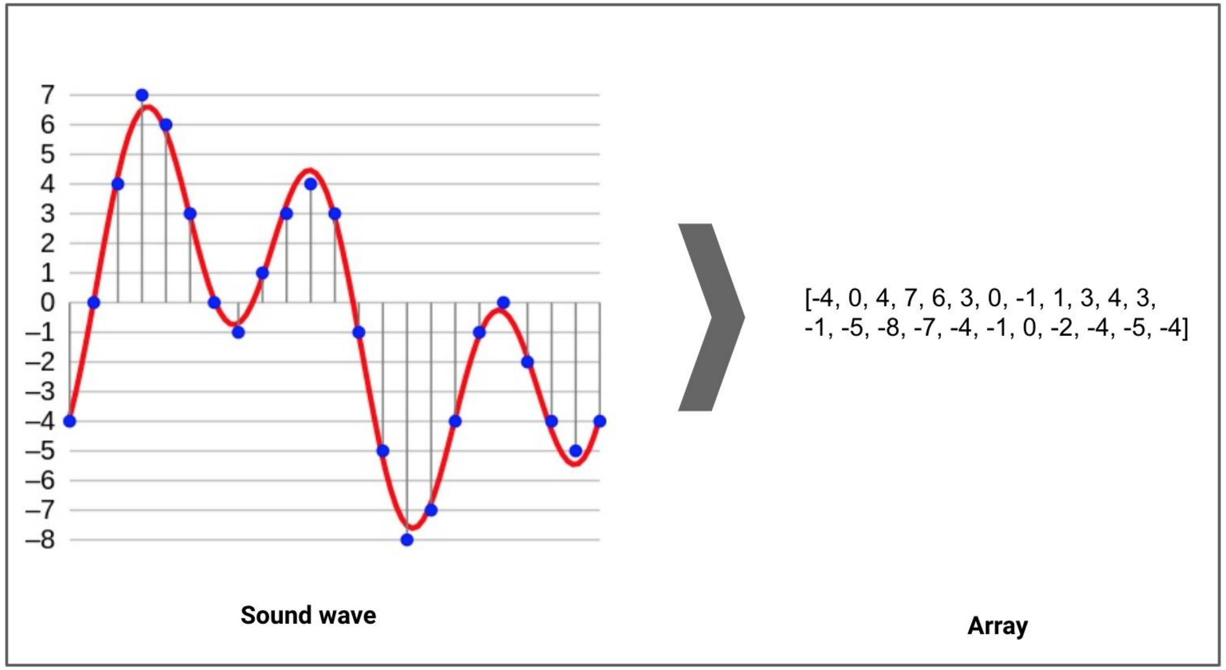
Music is nothing but a sound – Sound is nothing but a Wave

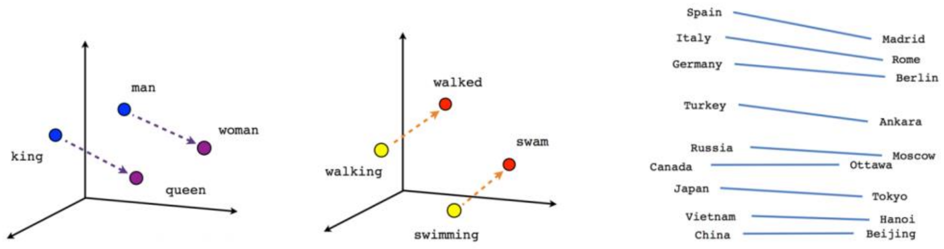
Wave consists of Amplitude , Time and Frequency – Either in Time domai

It is easy for us to represent and extract the features (Our DSP days – Not
Processing!!!)









Male-Female

Verb tense

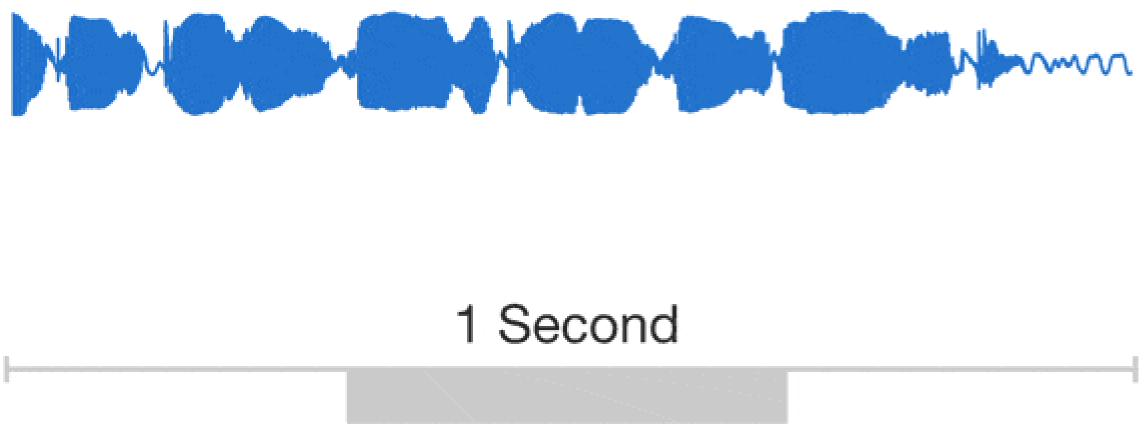
Country-Capital

How does Machine see Music?

Audio is an extremely rich data source. Depending on the sample rate — the number of points sampled per second to quantify the signal — one second of data could contain thousands of points

Wave Forms:
Waves are repeated signals that oscillate and vary in amplitude, depending on their complexity. In the real world, waves are continuous and mechanical — which is quite different from computers being discrete and digital.

So, how do we translate something continuous and mechanical into something that is discrete and digital?



Deep learning and Music Generation – Overall Training Process