BABES-BOLYAI UNIVERSITY CLUJ-NAPOCA

FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

SPECIALIZATION COMPUTER SCIENCE

DIPLOMA THESIS

Music and Feelings: A deep learning approach to emotional composition

Supervisor:
Darabant Sergiu Adrian

<div align="right">

Author:
Naiman Alexandru Nicolae

</div>

Cluj-Napoca
2020

WIP: Abstract

Year after year, the music industry sets new record on their revenue earning billions of dollars through different mediums, from LP records to CDs, and with the rise of the Internet, through streaming platforms. This growth can be associated with the ease of producing and releasing new songs, processes that have been facilitated through different tools such as Digital Audio Workstations and Virtual Studio Technology plugins. Even with all these advancements, the approach one has when composing did not change in the last hundreds of years: composing based on inspiration with a combination of trial and error corrected by music theory.

This thesis introduces a tool designed for easing music composition, flattening the steep learning curve of music theory, by helping to compose songs based on feelings. The main goal of the application is providing the user with a way of generating songs starting on some emotion given data. The user provides the input emotion values, and the software should respond with a composed song in which we can recognise the given emotion.

The application consists of two parts: the frontend web application with which the user interacts, and the backend on which our machine learning model is running. For the emotion data, we used the valence-arousal classification, which specifies that any feeling can be expressed through the two aforementioned values. For the composing algorithm, we used autoencoders and their capabilities of learning the internal structure of our data and used it for generating new songs based on new input.

This work is the result of my activity. I have neither given nor received unauthorized assistance on this work.

WIP: Abstract

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Music is a moral law. It gives a soul to the Universe, wings to the mind, flight to the imagination, a charm to sadness, gaiety and life to everything. It is the essence of order, and leads to all that is good and just and beautiful." [18]

1.1   Motivation

Music is a collection of coordinated sounds [10] whose medium is sound, being one of the first ways of expressing the human had ever had. From childhood to late adult life, we find ourselves listening to music, according to our feelings: from babies' lullabies to calm our little kids, to the most hyper energetic metal or drum and bass to keep us awake when learning for the finals.

The art of composing music is not one that any beginner can master due to the fact that music theory has a very steep learning curve. You can still compose music, without knowing anything about theory, but it will be very hard to know how to really express your thoughts since the notes do not have any meaning to you other than their sound. When you start learning, you will find out about scales, modes, chords progression and how they clearly relate to some feelings.

1.2   Solution Overview

In the latest years, with the technological advancements regarding artificial intelligence, dreaming about diagnosing a patient in seconds, facial recognition, instantaneous translating, and self-driving cars is no longer an idea in Asimov's mind. Nowadays, with a bit of work, one can create a neural network, train it on a given dataset and see the results about its objective, without needing to worry about the internal structure of each component inside the network.

I want to encourage and help all people that have difficulties composing music with the help of machine learning. I want to create an application that can compose, enhance or continue a piece of music based on some emotions that the user can provide. The music should have a structure and the app will generate a MIDI file containing a song with the given characteristics.

1.3   Thesis Outline

WIP: TODO: add thesis structure when done

Chapter 2

Machine Learning and Autoencoders

This chapter contains theoretical information about the paper's domain, deep learning, and the principal model used, autoencoders. Of course, before starting looking in its domain, we should start with the basics, namely, machine learning.

2.1 What is Machine Learning?

Machine learning is one of many subfields of artificial intelligence, the science of getting the computer to learn from experience without being explicitly programmed to do so, improving their ability to think, plan, decide, etc. [13].

Algorithms in this area are different in their approach, the type of data they input and output, and the type of problem they are learning to solve, but, despite all differences, they are based on the same idea that there are some generic algorithms that can discover patterns, without having to write code; instead, they build their own logic on the data fed to them.

2.2 Machine Learning Styles

When studying machine learning, one can identify a plethora of algorithms, each unique in its approach and purpose, but all of them can be classified based on the learning style. Therefore, there can be grouped into four main categories:

- **Supervised Learning**

- **Unsupervised Learning**

- **Semi-supervised Learning**

- **Reinforcement Learning**

2.2.1 Supervised Learning

In supervised learning, given a set of example input-output pairs, the job of the algorithms is to approximate a function that maps from input to output [14]. It is called supervised learning because human experts act as the teacher, where they serve the computer with training data containing the input and also provide the correct output, from which the computer should be able to observe patterns. As depicted in Figure 2.1, the algorithm should be able to predict the output when given new inputs, based on the function that it approximates [6].
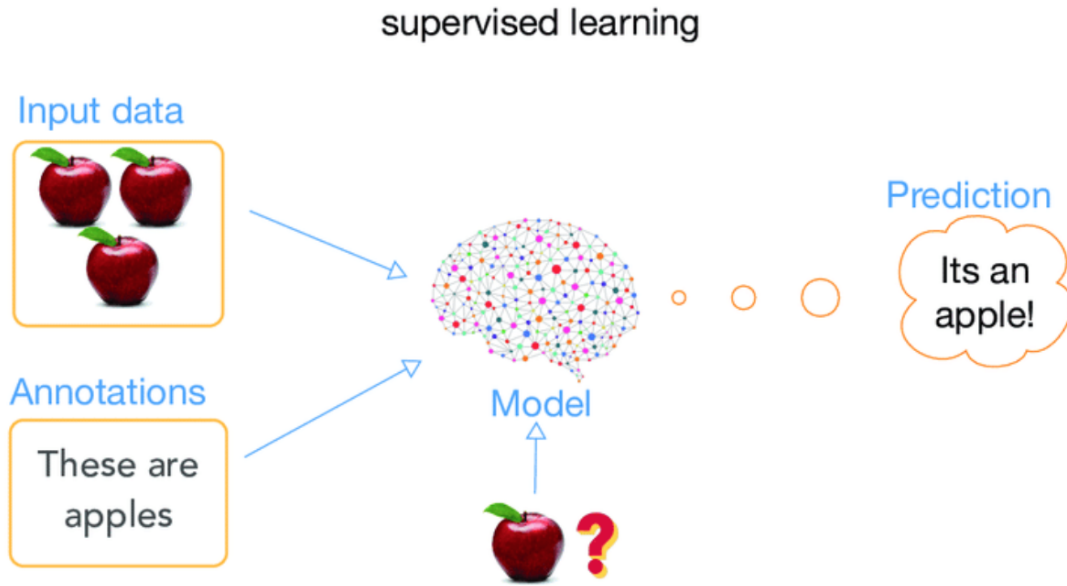

Figure 2.1: *Supervised algorithm predicting new output data given new input [19]*

Furthermore, supervised machine learning algorithms can also be grouped based on the type of problem they intend to solve. Having said that, there are classification and regression algorithms.

Classification algorithms solve problems in which the input data has been labelled in different classes or categories, meaning that the goal is to predict discrete values such as $\{0, 1\}$, $\{2, 4, 6, 8 \dots\}$, $\{cat, dog\}$, $\{spam, normal\}$, etc [6].

Regression algorithms, on the other hand, solve problems in which the output variable is a real value, meaning that the goal is to predict continuous values based on the input data, such as predicting house prices based on their size and location [6].
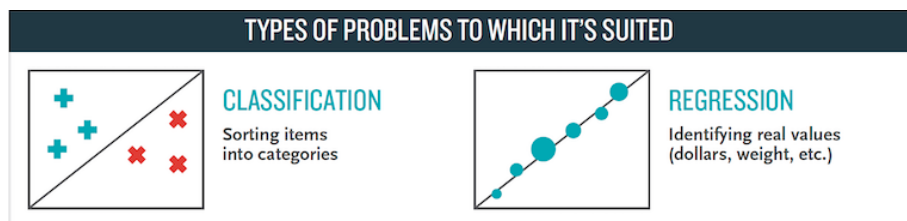

Figure 2.2: *Classification(left) and Regression(right) visualization [2]*

2.2.2 Unsupervised Learning

In unsupervised learning, the job of the algorithms is to observe patterns in the structure of the given data even though no explicit feedback is supplied [14]. There is no human factor implied in the learning process, hence the name of unsupervised learning. As shown in Figure *2.3*, they should be able to identify the underlying structure in the given data, to better understand it [2].
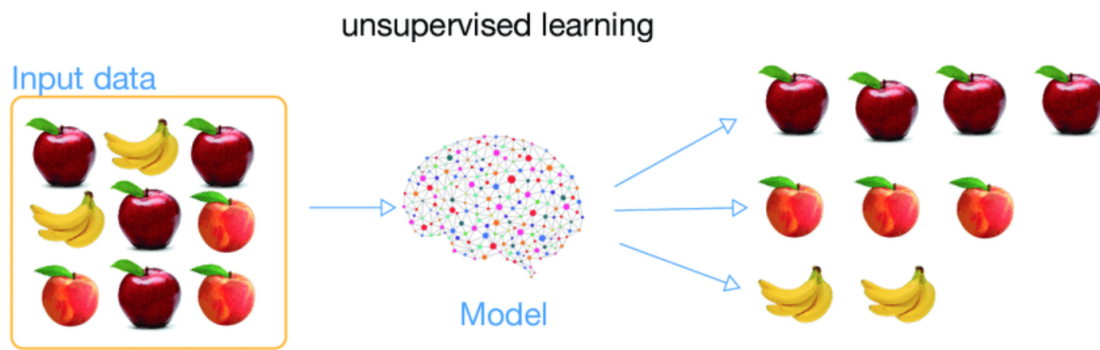

Figure 2.3: *Unsupervised algorithm grouping data by their observed features [19]*

The same as supervised learning, unsupervised algorithms can also be grouped based on the type of problem they intend to solve. Provided that, there are clustering and low-dimensional embedding algorithms [14].

Clustering algorithms try to solve problems by grouping the input data into different segments based on similarities learned, such as grouping customers by purchasing behaviour [1].

Low-dimensional embedding, also called dimensionality reduction, is a technique in which complex data, that is difficult to describe (data that needs more than two or three dimensions to represent) is reduced to a lower number of dimensions, wherein the machine learning algorithms can help since they can learn the internal structure inside our data, furthermore, transcribing it to an easier way of interpreting it [14].

2.2.3 Semi-supervised Learning

Semi-supervised learning algorithms lie between the aforementioned categories. They intend to solve problems where you have a large amount of data but only some labelled. A lot of machine learning problems are using this technique for the reason that labelled data can be expensive or time-consuming to obtain, in contrast to unlabelled data that is much easier to acquire [1].

As shown in Figure *2.4*, one can use an initial classifier on the labelled data, classify based on the features learned previously, the unlabelled data, retrain the classifier with the whole dataset, then use it to predict new outputs given new input data, ideally obtaining a better model [9]
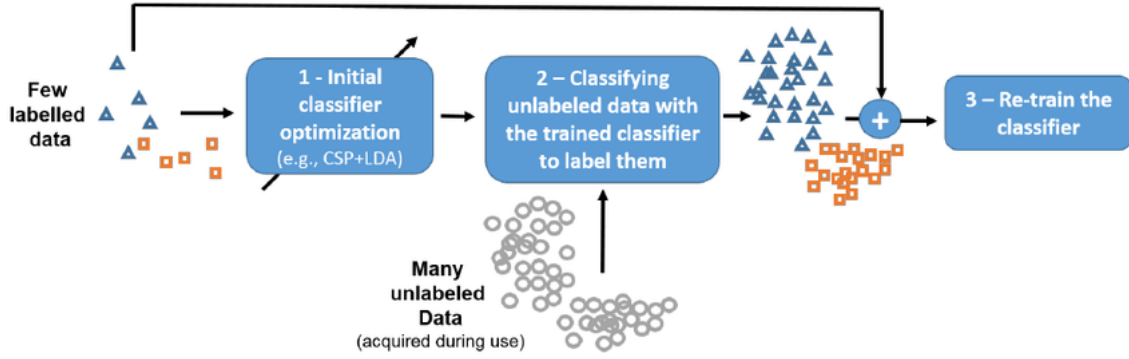
Figure 2.4: *Semi-supervised approach on data classification [9]*

2.2.4 Reinforcement Learning

Reinforcement learning algorithms are based on a system of type reward-punishment, where the job of the algorithm is to iteratively interact with its environment, making actions in such a manner that will maximize the rewards or minimize the punishments. To put it another way, in reinforcement learning, the software learns from its past experiences allowing it to develop an optimal behaviour within a specific setting, where it aims to increase its performance [6].
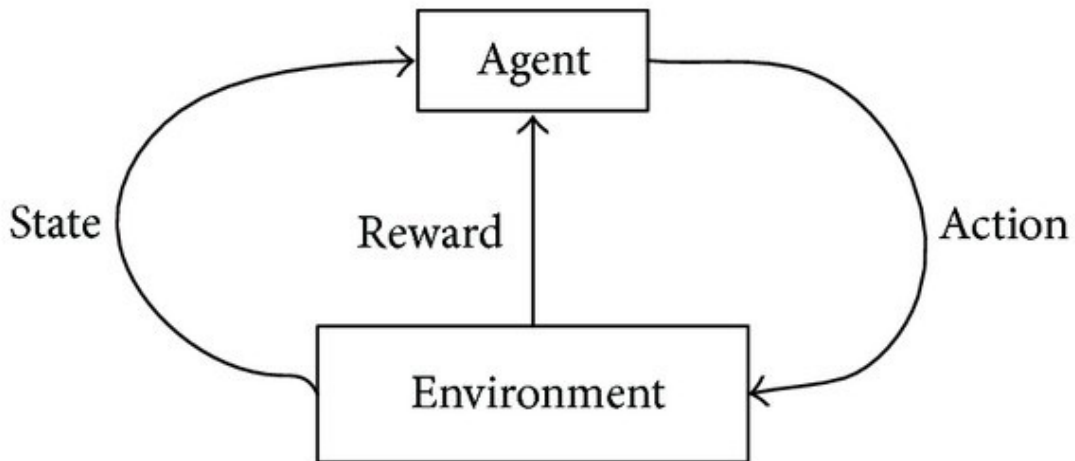
Figure 2.5: *Scheme of how reinforcement algorithms work [6]*

Figure *2.5* is a visual representation of how reinforcement learning works. The algorithm (called the "agent" in the picture), takes action in the environment, which will possibly trigger a reward. After that, the state is updated, and when deciding its next action, it will take into account its last state and the result of the previous action.

2.3 ANNs and how they work

One of the main types of machine learning algorithms that are vastly used is artificial neural networks or ANNs. They are structured as weighted graphs, modelled after the human brain where each node represents a neuron and each edge represents the

synapses of the neural network. The weight of each edge determines how powerful the synapse is. The neurons are distributed in groups named layers. Neurons can form synapses only with neurons from other layers [20].

2.3.1 Artificial neurons and activation function

The smallest unit of an artificial neural network is the **artificial neuron**, namely the nodes of the graph. It is the place where the computation happens: a node merges the input data by doing their weighted sum, to dampen or amplify the input. Further, the result will be given to a so-called **activation function** whose purpose is to determine whether or not the neural signal will be passed on to the next neurons or not [5]. A more detailed explanation of the subject can be seen in Figure *2.6*.
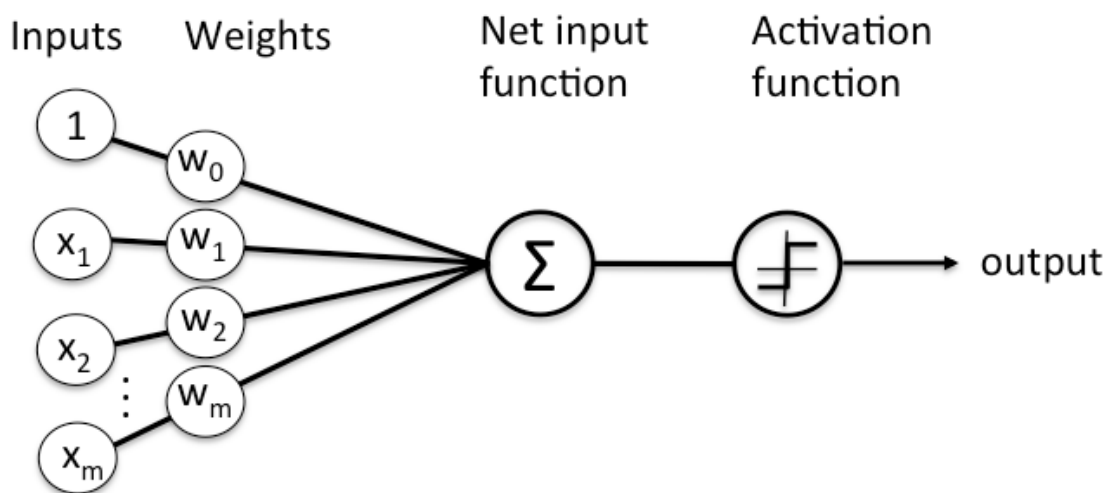

Figure 2.6: *Anatomy of an artificial neural network [5]*

The aforesaid **activation function** is a mathematical expression that helps the ANN to add non-linearity to itself. With more and bigger datasets, the patterns that need to be identified become more complex. With the help of these types of function, the input is much easier to analyze and interpret.

- **zero-centered**: output should be symmetrical at zero so that it does not move to a particular direction;

- **computationally inexpensive**: the function is called for every layer for each input data, meaning that it could be invoked thousands of times. Therefore, they should not be computationally expensive;

- **differentiable**: at the core of almost every machine algorithm lies an optimization algorithm. One of the most popular approaches is the gradient descent algorithm in which one of the most essential features is that the function used needs to be differentiable; this idea will be discussed in detail in a subsequent paragraph;

2.3.2 Backpropagation and gradient descent

Backpropagation describes how machine algorithms learn. As shown in Figure 2.7, after the data traverses the neural network in a forward manner, propagating the signals through the neurons, the algorithm will go in reverse adjusting each weight for each node and its links [7]. This technique is called backpropagation of errors, or simply, backpropagation and is the learning engine of the most ANN's.
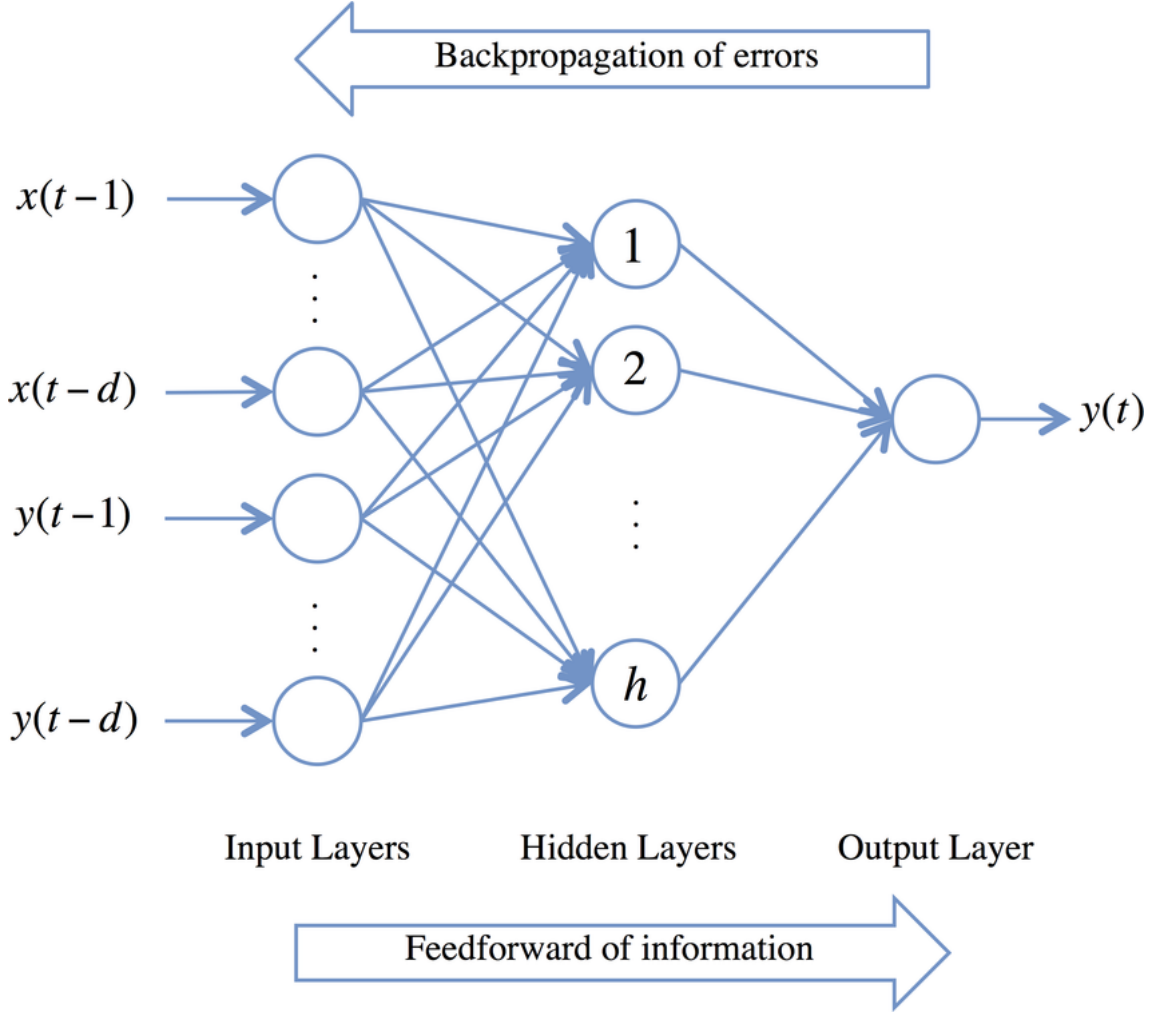

Figure 2.7: *Feedforward Backpropagation Neural Network architecture. [3]*

The **gradient descent**, as mentioned before, is an optimization algorithm best-used to find the (local) minimum value for a function [1]. Hence, it can be used to reduce the errors in our algorithm, since the training of a neural network is just the process of finding the set of weights and other parameters in such a fashion that the errors(differences) between our expected and actual results are minimum. Hence, the gradient descent is used to compute the errors of our algorithm, that will be later backpropagated through the ANN.

2.3.3 Layers of ANNs

Multiple artificial neurons are grouped in **layers**; the purpose of each one is to apply a non-linear transformation of the input from one vector space to another [8]. As

you can see in *2.8* there are three types of layers in an artificial neural network:

- Input layer

    This layer of a neural network is the very beginning of the ANN's workflow, bringing the initial data into the system for further processing by subsequent layers of artificial neurons [16].

- Hidden layer

    Any layer between the output and the input layer is considered to be a hidden layer. The number of neurons they contain and also their number can vary. Their role is to take in a set of weighted inputs and produce an output through an activation function [15].

- Output layer

    This layer of a neural network is the last layer of neurons that produces given outputs for the program. They usually are made much like any other artificial neuron, but they may be observed in a different way, the output layer coalesces and concretely produces the end result [17].
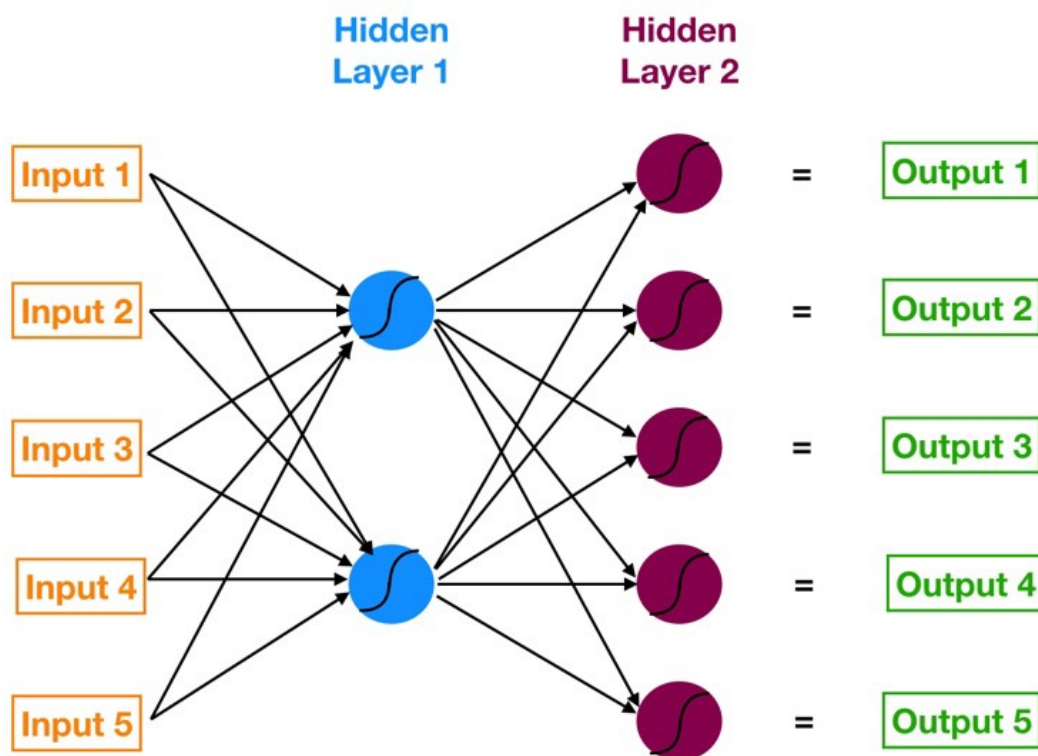
Figure 2.8: *Neural network with one hidden layer [20]*

2.4   What is Deep Learning?

Deep learning is a subfield of machine learning that contains a collection of ANNs that are known for their capability on learning unsupervised from data that is un-

structured or unlabeled, also known as deep neural learning or deep neural network.

A second classification we can make on ANNs is based on the relationships between their nodes. Then, neural networks can be recurrent or feedforward; the first one does not have any loops in its graph and can be organized in layers. A deep neural network is a feedforward ANN with many hidden layers.

A visual representation of the aforementioned is the Figure *2.9* wherein we could see the differences between a simple feedforward ANN with one hidden layer(left) and a deep neural network with three hidden layers.
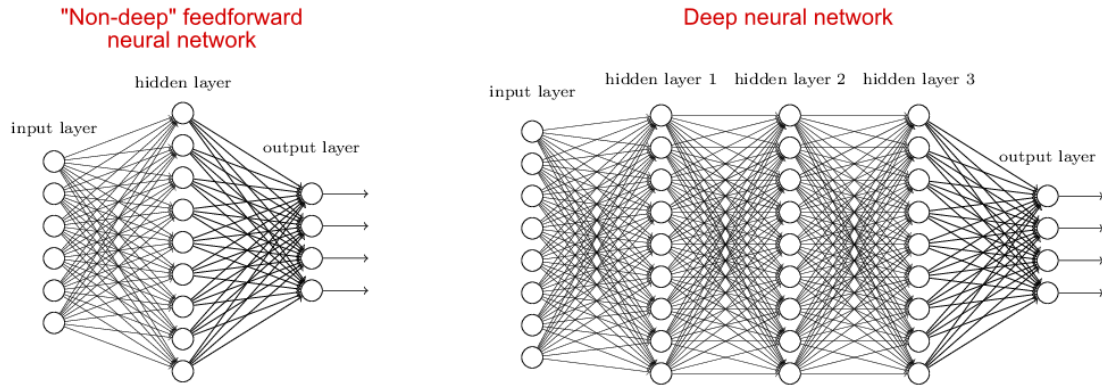

Figure 2.9: *Difference between a Non-deep ANN(left) and a deep ANN(right) [11]*
.

Deep learning is an area in machine learning that achieves state-of-the-art results by learning to represent the world "as a nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones".[7]

A visual representation of the aforementioned statement can be seen in Figure 2.3.

2.5   Autoencoders and how they learn

Autoencoders are neural networks usually used for dimensionality reduction. They learn in an unsupervised manner, their purpose being to obtain the internal representation of the data by compressing and then reconstructing the original input, making a copy of it.[11]

This type of neural network is composed of two parts:

- **encoder**
- **decoder**

The first one may be viewed as a network that aims to compress our data into a lower dimension code, named encoding. It could be represented as a function $\boldsymbol{h} = f(x)$.[11]

The second one intends to reconstruct the input based on that encoding, removing thus all the possible noise. It could be represented as a function $r = g(h)$.[11]
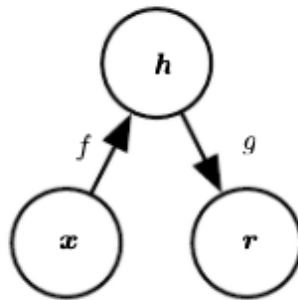
Figure 2.10: *Structure of an autoencoder [11]*

.

Therefore, as shown in Figure 2.10, the entire autoencoder can be seen as the following function: $g(f(x)) = x$. In other words, an autoencoder tries to mimic the identity function, but, due to the compression, is unable to do so. However, thanks to that same compression, it learns how to successfully identify the most important features of our data's internal structure, represented by the encoding.

Chapter 3

Music and MachineLearning

WIP: small notes: - furture work - Lakhness -¿ care doar genereaza muzica - aiva.com -¿ genereaza muzica based on genre, de unde poti corela emotii, dar nu emotii in mod direct - astia vorbesc de cum ar trebui sa arate un astfel de tool -¿ pot introduce diagrama lor ca referinta si sa descriu eu flow-ul singur -¿ pot face usor tranzitie la proiectul meu

Chapter 4

Emotion MIDI Dataset

WIP

Chapter 5

Proposed approach

5.1 Problem definition

With the advancements of technology, people tried digitalizing every aspect of their life. They renounced the traditional slow postal services in favour of the modern fast email. Digital photos gained popularity to the detriment of physical copies. In the same way, CD-s and other physical mediums were forgotten with arising of digital MP3s.

Over the past years, with the rise of the internet, from digitalizing their lives, people started to sync it with the cloud. Hence, a new way of storing and transmitting data emerged: streaming.
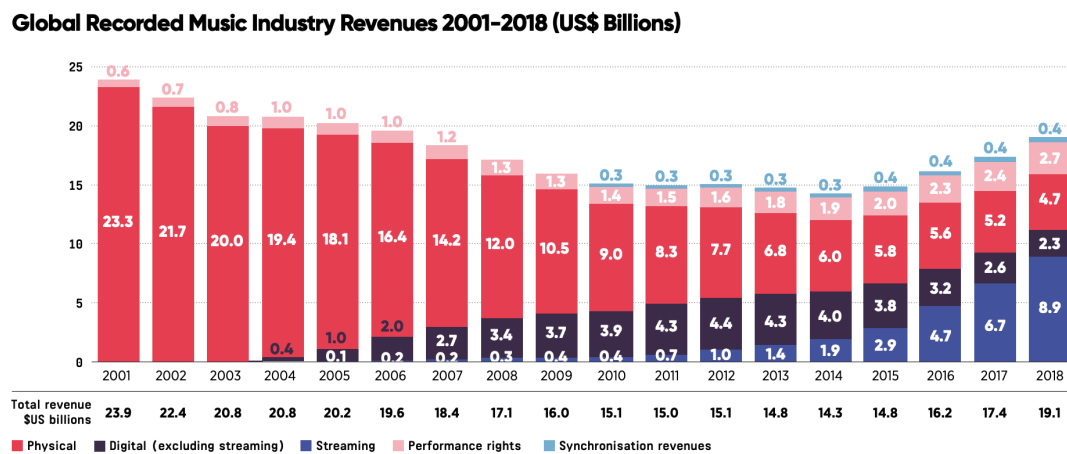
Global Recorded Music Industry Revenues 2001-2018 (US$ Billions)

| Total revenue $US billions | 23.9 | 22.4 | 20.8 | 20.8 | 20.2 | 19.6 | 18.4 | 17.1 | 16.0 | 15.1 | 15.0 | 15.1 | 14.8 | 14.3 | 14.8 | 16.2 | 17.4 | 19.1 |

■ Physical ■ Digital (excluding streaming) ■ Streaming ■ Performance rights ■ Synchronisation revenues

Figure 5.1: *Global Recorded Music Industry Revenues 2001-2018 [12]*

As depicted in Figure *5.1*, in the last decade, the revenue of the musing industry shifted, going from physical to digital, in the end, the highest revenue being generated by streaming means.

With this change, one can easily compose and produce music from their studio or even bedroom, since it is not compulsory to be tied to a record label. There are lots

of successful independent musicians who create music in this way and the numbers are growing more and more every year [4].

Even with all these developments, the act of composing music did not drastically change in the last few hundred years. If one wants to compose a song, they should learn how to use instruments, and how to put notes in such a manner that they sound good together. Technology eased the process of composing, but one must still learn chords, scales, and other elements of music theory to make the process less hazardous. Although music theory has a steep learning curve, learning it helps express oneself musically. Despite all these advancements, there is a lack of tools facilitating the creative process.

5.2   Functional requirements

This section contains details about the functional requirements of the application and its main purpose, composing music based on feelings.

The application software allows the user to input arousal and valence values and to tweak other features that our model learned, and it should generate a song.

Bellow, there is a list of the main features the app has:

- The user should be able to input values for valence and arousal and the app should respond with a composed song based on them

- The user should be able to listen to the generated song

- The user should be able to change the instrument in which the song is rendered

- It should be capable of modifying the input and see live edits of the displayed song

- The app should allow the user to export the file as MIDI and save it to its computer

Chapter 6

Application

WIP

Chapter 7

Conclusion and future work

WIP

Bibliography

[1] J. Brownlee. *Master Machine Learning Algorithms: Discover How They Work and Implement Them From Scratch.* Jason Brownlee, 2016.

[2] Okan Bulut and Christopher Desjardins. *Exploring, Visualizing, and Modeling Big Data with R.*
https://okanbulut.github.io/bigdata/.

[3] Jacinta Chan Phooi M'ng and Mohammadali Mehralizadeh. Forecasting east asian indices futures via a novel hybrid of wavelet-pca denoising and artificial neural network models. *PLOS ONE*, 11:e0156338, 06 2016.

[4] M. Daniels. *Why Independent Musicians Are Becoming The Future Of The Music Industry.*
https://www.forbes.com/sites/melissamdaniels/2019/07/10/
for-independent-musicians-goingyour-own-way-is-finally-starting-to-pay-off/
#27f0c8a314f2.

[5] Arden Dertat. *A Beginner's Guide to Neural Networks and Deep Learning.*
https://pathmind.com/wiki/neural-network.

[6] David Fumo. *Types of Machine Learning Algorithms You Should Know.*
https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-k

[7] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016. http://www.deeplearningbook.org.

[8] Vandit Jain. *Applied Deep Learning - Part 1: Artificial Neural Networks.*
https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-ne

[9] Fabien Lotte. *Signal Processing Approaches to Minimize or Suppress Calibration Time in Oscillatory Activity-Based Brain–Computer Interfaces.* *Proceedings of the IEEE*, 103:871–890, 06 2015.

[10] Samer Maini, V. Sabri. *Machine Learning for humans.*
https://medium.com/machine-learning-for-humans/
why-machine-learning-matters-6164faf1df12/.

[11] M.A. Nielsen. *Neural Networks and Deep Learning.*
http://neuralnetworksanddeeplearning.com/.

[12] International Federation of the Phonographic Industry. *Global Music Report 2019: State of the Industry Report.*

[13] Stephanie Przybyl. *What is Music? - Definition, Terminology & Characteristics*.
`https://study.com/academy/lesson/what-is-music-definition-terminology-characte`
`html`.

[14] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*.
Prentice Hall Press, USA, 3rd edition, 2009.

[15] Technopedia. *Hidden Layer*.
`https://www.techopedia.com/definition/33264/`
`hidden-layer-neural-networks`.

[16] Technopedia. *Input Layer*.
`https://www.techopedia.com/definition/33262/`
`input-layer-neural-networks`.

[17] Technopedia. *Output Layer*.
`https://www.techopedia.com/definition/33263/`
`output-layer-neural-networks`.

[18] Watson D. *THE WORDSWORTH DICTIONARY OF MUSICAL QUOTATIONS*. Wordsworth Reference, Edinburgh, 1991.

[19] Ma Yan, Kang Liu, Zhibin Guan, Xinkai Xu, Xu Qian, and Hong Bao.
Background Augmentation Generative Adversarial Networks (BAGANs): Effective Data Generation Based on GAN-Augmented 3D Synthesizing. *Symmetry*, 10:734, 2018.

[20] Tony Yiu. *Understanding Neural Networks*.
`https://towardsdatascience.com/understanding-neural-networks-19020b758230`.