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

Author:
Naiman Alexandru Nicolae

Cluj-Napoca
2020

Thesis Title

Thesis Subtitle

Author Name

Abstract

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." [12]

1.1 Motivation

Music is a collection of coordinated sounds [7] 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. [5].

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 [8]. 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 seen depicted in the Figure *2.1*, the algorithm should be able to predict the output given new inputs, based on the function that it approximates [3].

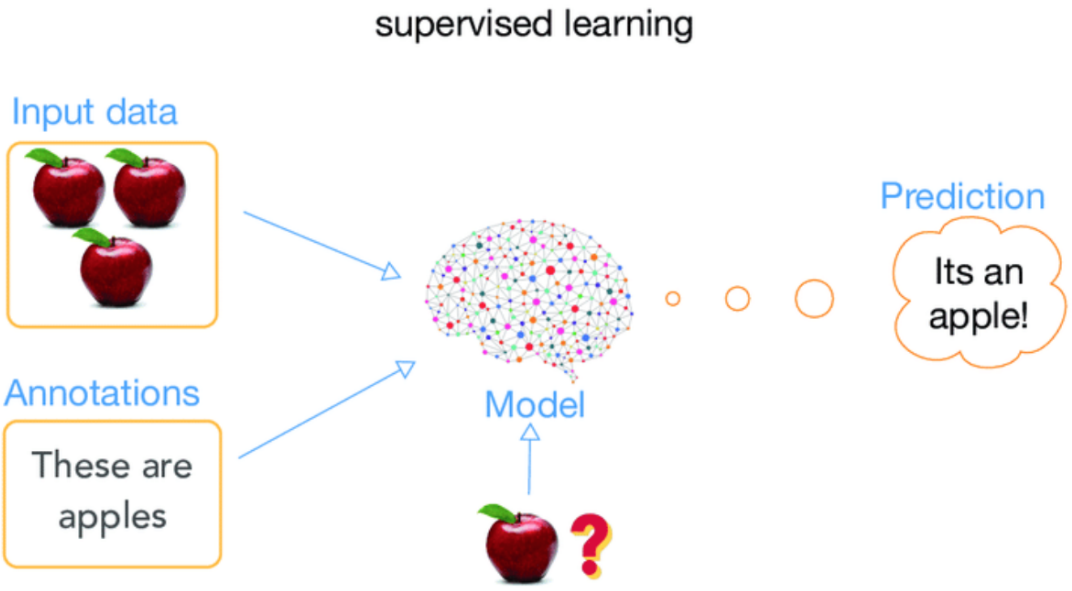

Figure 2.1: *Supervised algorithm predicting new output data given new input [13]*

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 $\{0, 1\}$, $\{2, 4, 6, 8 \ldots\}$, $\{cat, dog\}$, $\{spam, normal\}$, etc [3].

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 [3].

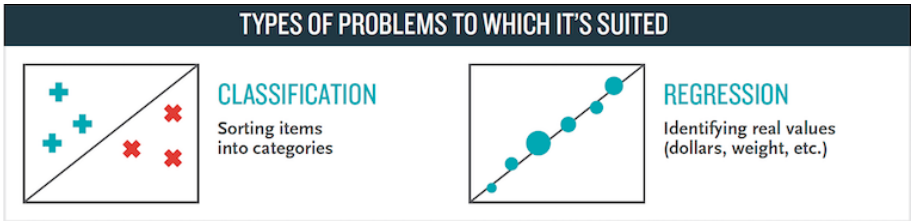

Figure 2.2: *Classification(left) and Regression(right) visualization [2]*

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 [8]. 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].

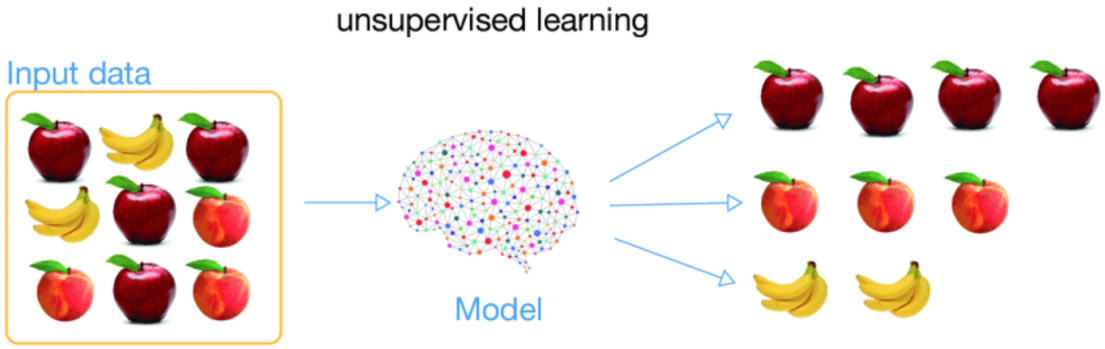

Figure 2.3: *Unsupervised algorithm grouping data by their observed features [13]*

Just 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 [8].

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 [8].

2.2.3 Semi-supervised Learning

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 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 [4]

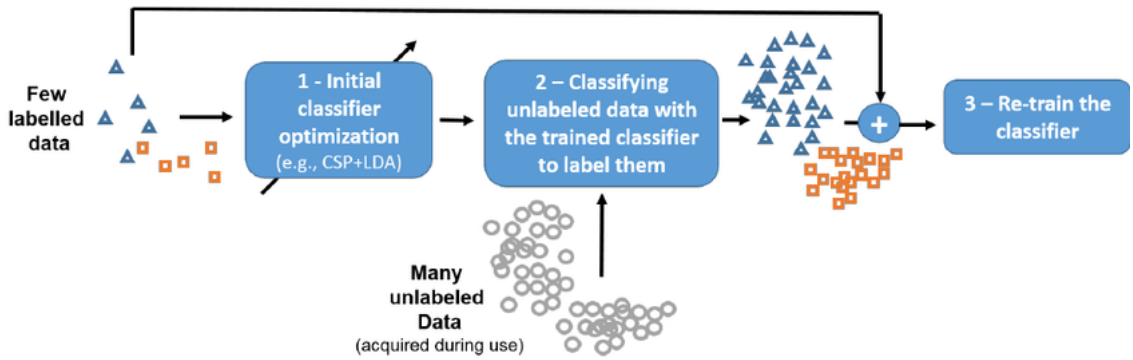

Figure 2.4: *Semi-supervised approach on data classification [4]*

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.

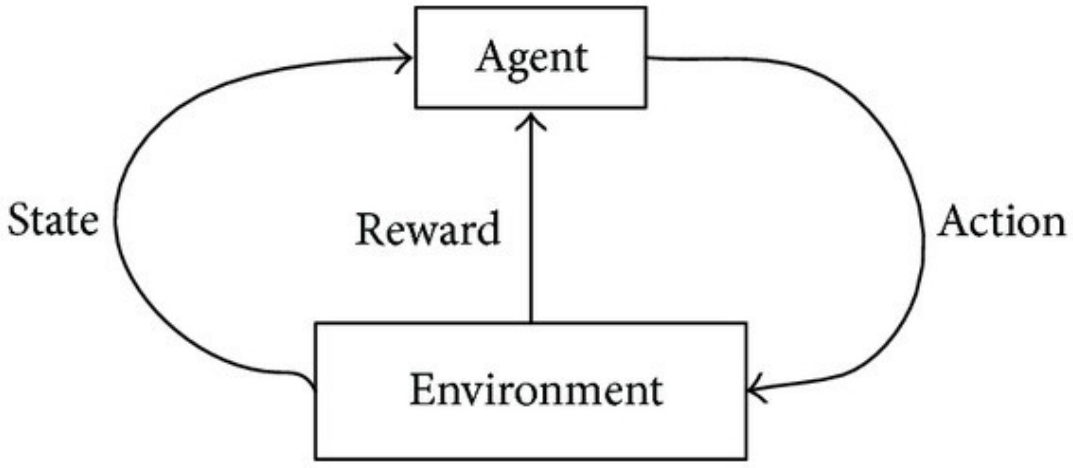

Figure 2.5: *Scheme of how reinforcement algorithms work [3]*

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 ANN's 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.

2.3.1 Artificial neurons, activation function and backpropagation

As you can see in *2.6* 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 [10].

- 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 [9].

- 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 [11].

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 unstructured 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.

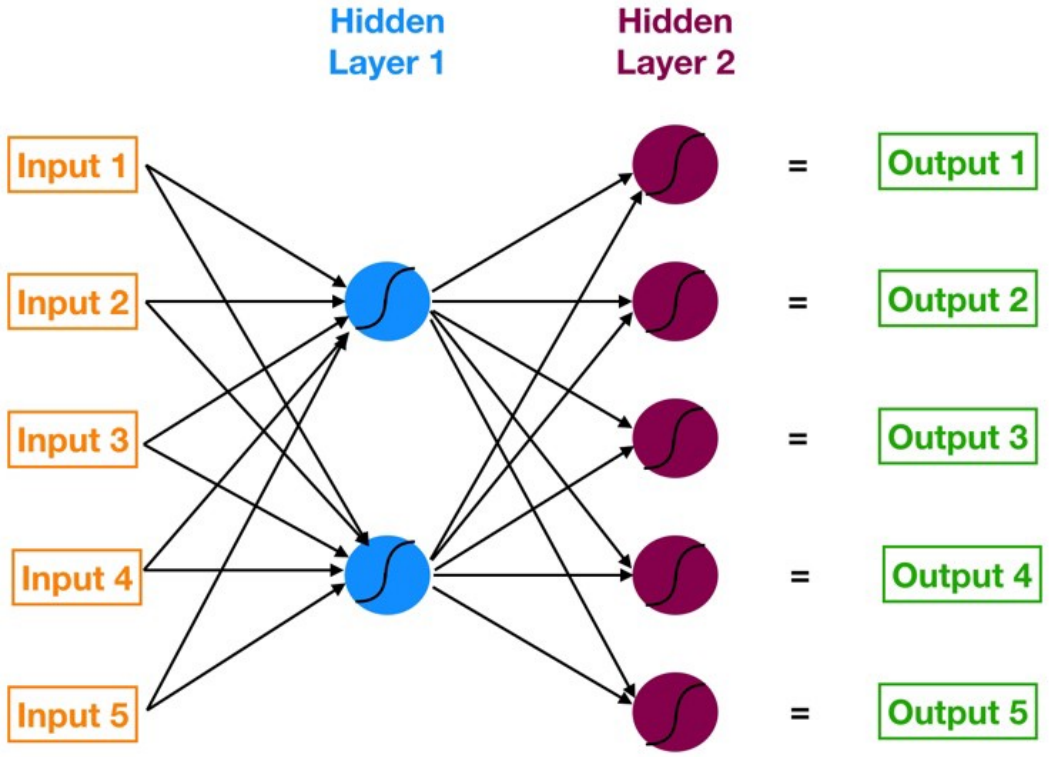

Figure 2.6: *Neural network with one hidden layer [14]*

A visual representation of the aforementioned is the Figure *2.7* 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

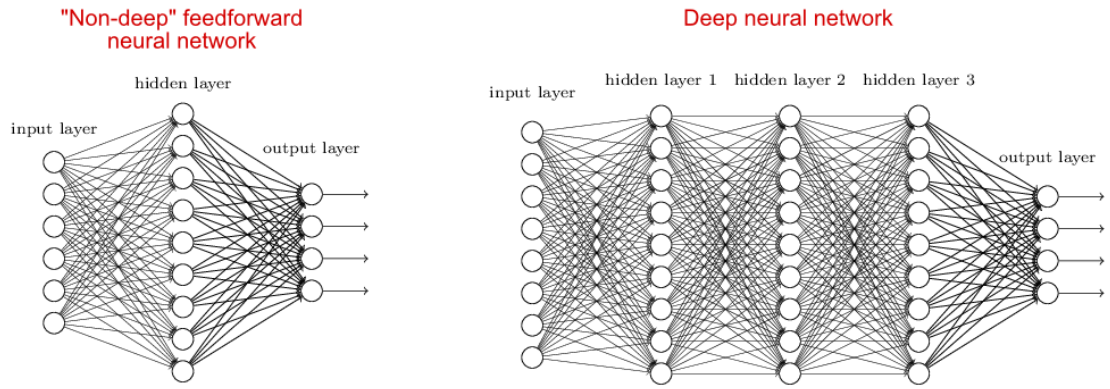

Figure 2.7: *Difference between a Non-deep ANN(left) and a deep ANN(right) [6]*

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

[4] 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.

[5] Samer Maini, V. Sabri. *What is Music? - Definition, Terminology & Characteristics*.
http://neuralnetworksanddeeplearning.com/.

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

[7] Stephanie Przybyl. *Machine Learning for humans*.
https://medium.com/machine-learning-for-humans/
why-machine-learning-matters-6164faf1df12/.

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

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

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

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

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

[13] 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.

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