

Student: Jason Zhou

Mentor: Dr. Dongjin Song

Status Report #: 1

Time Spent on Research This Week: 2.5 hours (9/5 - 9/12)

Cumulative Time Spent on Research: 7 hours

Miles Traveled to/from Mentor This Week: 0

Cumulative Miles Traveled to/from Mentor: 0

=====

Wednesday, August 25th, 2021 (1 hour):

For the first day of research, I spent the day learning the basics of machine learning. To start, I watched a quick course on how to get ready and downloaded a software called Jupyter. From my understanding, this is an IDE (Integrated Development Environment) that is ideal for coding that is related to data science and machine learning. Afterward, I ran through a couple of exercises that the video course gave. These exercises focused on basic commands that are useful in data science and machine learning. I took notes on the various commands and vocabulary so that I could review them when I needed to. Although the video is around 50 minutes, I went through it quite slowly to make sure I had a good understanding of everything. After reaching the halfway point of the video, I decided to continue this video the next day.

Thursday, August 26th, 2021 (1 hour):

On the second day, I continued to watch the video I had started the previous day. This day, I learned the steps to make a machine learning program, which is labeled on the picture I have added below. Also, I actually used a machine learning algorithm and created a program that learned to predict an outcome based on a pattern. I used a very small dataset that had people's ages and what music those people liked. Using this data, I predicted what type of music other people outside this data might like. Although this program was extremely basic, it helped me gain valuable experience on how to create machine learning programs. Despite this, I still lacked knowledge in many areas. After finishing the video, I decided that I would start with an overview of machine learning and the different algorithms that exist within the field.

```
import pandas as pd
from sklearn.tree import DecisionTreeClassifier

music_data = pd.read_csv('music.csv')
X = music_data.drop(columns=['genre'])
y = music_data['genre']

model = DecisionTreeClassifier()
model.fit(X, y)

predictions = model.predict([[21, 1]])
```

import libraries
import data set
create model
predict model
predict something

(Picture of Some Code That I Wrote. A very simple illustration of the steps required to make a machine learning algorithm.)

Friday, August 27th, 2021 (1 hour):

On this day, I watched a video on an overview of the various aspects of machine learning such as the different algorithms that exist. Specifically, there are essentially two different categories of algorithms that one can use in machine learning: regression and classification. In extremely simple terms, Regression could be likened to linear regression (predicting patterns based on given data points; continuous results). Classification has to do with discrete data. For example, I could say that biology is a field of science. This would count as a form of classification. Additionally, besides algorithms, there are two different ways to learn: supervised and unsupervised learning. Supervised learning is when you give the machine a set of inputs and relate them to a set of outputs. In contrast, unsupervised learning, from what I understand, is when you give the machine a list of inputs without corresponding outputs. In both cases, the program will have to identify a pattern in the data in order to make predictions.

Friday, September 3rd (1.5 hours):

I decided to pick off where I left off last week and learn more about what I don't know. Again, I searched up a video on Youtube to find courses on machine learning. As opposed to the last video I watched, this one was ten hours long, so it goes more in-depth. Although the beginning parts were similar to the last video I watched, I decided to not skip the start because it served as a good review (and it might fill in anything I could have missed). For example, previously, I thought that there were two ways of learning: supervised and unsupervised learning. However, there is actually a third type, which is called reinforcement learning. This type of learning is reminiscent of how humans learn. Specifically, this type of learning focuses on learning from past mistakes and slowly improving as time goes on. As the video dragged on for a long time, this was all I got to. Overall, it was a good review, which also allowed me to obtain new knowledge.

Tuesday, September 7th (10 minutes):

After being in school for a little more than a week, I have gained a better understanding of how ARM will operate and what the schedule for the year will more or less look like. As such, I decided to take to this day to email my mentor, Dr. Dongjin Song about what I had learned about the course. In summary, I told him that I would begin sending weekly status reports, the research proposal that is due at the end of the first quarter, the presentation I will need to give in the Winter, and that the Spring is when I will really get to dig into my research. Additionally, I also asked Dr. Dongjin Song if we could set up weekly Zoom meetings just to make communication a little easier. For example, these meetings would be a good time to ask any questions about machine learning that I had. If we have these meetings, I am hoping to have them every Friday after or around 4:30 (because I have cross country practice until this time).

Thursday, September 9th (1 hour):

Previously, my mentor and I had discussed that I might be working to develop a program (or AI) that can distinguish between environmental sounds. As such, I thought it would be useful to get a good understanding of what sound is and how it can be processed by computers. After searching on Youtube, I found a video series that covers the basics of sounds and waveforms but also goes in-depth on these topics. However, I was having trouble focusing, so I decided to continue at another time.

Also, I received an email from my mentor saying that he was able to have a meeting this Friday on Zoom. He also mentioned that he would like to organize training for a group of high schoolers (including me) to get familiar with different topics and skills that we may need.

Friday, September 10th (15 minutes):

On this day, the Zoom meeting I had with my mentor revolved around the subject of how to proceed with ARM. Referencing his email from Thursday, he spoke about how he knew other colleagues who were mentoring high school students. As such, he suggested that we form a group. This would allow the other mentors to make a curriculum to outline what the students would need to learn. To do this, my mentor mentioned changing the ARM agreement because the students would have the opportunity to meet with Ph.D. students and possibly use UConn facilities.

Additionally, I asked if my mentor had any recommendations on what I should be studying. He advised me to look into libraries such as TensorFlow, SKLearn, and Numpy, which are all data libraries that are commonly used in machine learning (libraries are data packets that add new functionality and commands for certain tasks). I was aware of them before; however, I will make sure to develop an in-depth understanding of them.

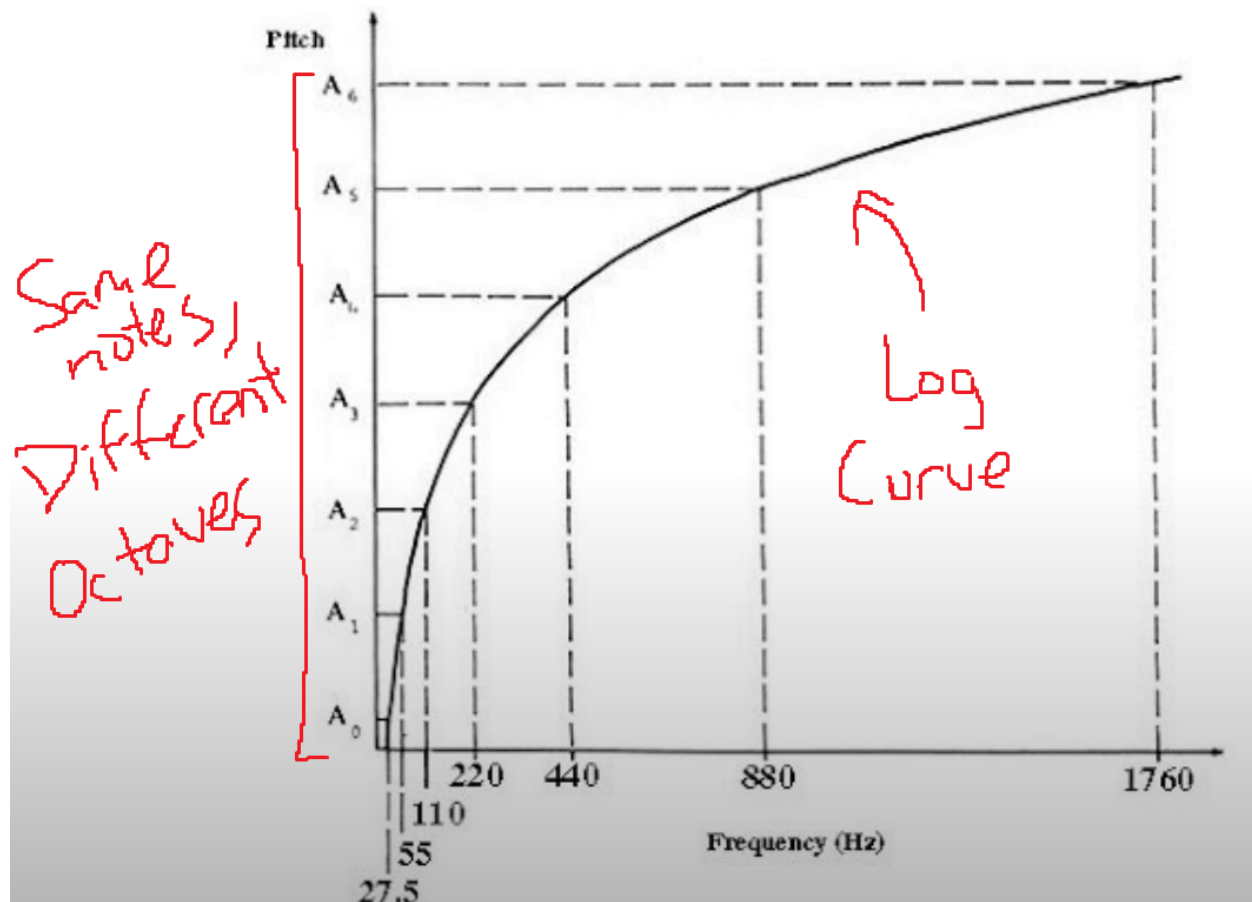
Finally, my mentor stated that he was thinking of hosting weekly Zoom meetings with other high school students, so it wouldn't only be me and my mentor in a Zoom meeting. I also believe that these students aren't limited to Glastonbury High School but may also come from different schools such as EOSmith.

Sunday, September 12th (1 hour):

I finally decided to continue the video I had been watching on Thursday. As the video was about sounds, most of it was review because I had learned it in physics last year. However, there were new aspects. For instance, there are generally four different types of waveforms (waveforms are the visualization of sound): simple, complex, continuous, and transient. Simple waveforms are like the standard sine or cosine curve. Complex waveforms are the results of multiple sine waves. Continuous waveforms are waves that do not show a pattern for periodicity and, as such, can be perceived as noise. Finally, transient waves are like sounds that are impactful but quickly fade away like a pop or boom. Besides the different types of waveforms,

the video also went over the relationship between octaves and pitch. For example, if one were to graph the same notes of different octaves on the y-axis and the frequency of each note on the x-axis, it would resemble a logarithmic curve.

Pitch-frequency chart



(Pitch-Frequency Chart. Shows the relationship between notes on different octaves and their corresponding frequencies.)

I will probably do more research into this Pitch-frequency chart later and try to find out what the exact importance of it is. However, for now, this is as far as I have gotten.

References

Hamedani, M. (2020, September 17). *Python machine learning tutorial (data science)* [Video].

Youtube. <https://www.youtube.com/watch?v=7eh4d6sabA0&t=631s>

Simplilearn. (2018, September 19). *Machine learning basics | what is machine learning? | introduction to machine learning | simplilearn* [Video]. Youtube.

<https://www.youtube.com/watch?v=ukzFI9rgwfU&t=366s>

Simplilearn. (2020, December 24). *Machine learning with python full course in 9 hours | machine learning tutorial | simplilearn* [Video]. Youtube.

<https://www.youtube.com/watch?v=eq7KF7JTinU>

Velardo, V. (2020, June 29). *Sound and waveforms* [Video]. Youtube.

<https://www.youtube.com/watch?v=bnHHVo3j124>