

ML Club Overview

ML Club Officers

October 2020

1 Introduction

TJ Machine Learning Club's purpose is to teach students about both the theory and practical applications of machine learning. Throughout the year, we'll go through the foundations of machine learning while incorporating coding competitions.

Until school reopens, all lectures will be recorded and posted on YouTube. Once our sponsors take attendance, we will send a link to the YouTube lectures for that day. You can ask questions regarding the lectures at anytime during the meeting.

1.1 Leadership and Contact Info

- Captain: Charlie Wu
- Captain: Saahith Janapati
- Teaching Coordinator: Caroline Sun
- Teaching Coordinator: Aarav Khanna
- Sponsors: Mr. Jurj and Mrs. Anderson

If you have any questions, you can reach out to any of us on Facebook or at tjmachinelearning@gmail.com.

2 What is Machine Learning?

Machine learning is a subset of Artificial Intelligence that attempts to generate models based off data in a non-explicit way. The goal of a model is to find patterns embedded within data and use what it learns to predict characteristics on unseen data.

2.1 Supervised Learning

Supervised learning algorithms analyze known training data with labels, the characteristics of the data that we want to predict, to predict the labels of unseen data. For example, an e-mail spam filtering algorithm would analyze previously seen e-mails that are already labeled as being spam or non-spam to predict whether new, unseen e-mails are spam or non-spam. A supervised learning problem where the class labels are discrete (i.e. are made up of distinct categories), such as the spam filter example, is called a classification task. Regression is another type of supervised task where the predicted value is continuous (e.g. predicting a student's SAT score based on their GPA).

2.2 Unsupervised Learning

Unsupervised learning algorithms analyze unlabeled training data. One common unsupervised learning task is clustering, which creates different groups for data and categorizes similar data into each group. It seeks to determine how data is organized without labels on each data point. An example might be categorizing visitors of a website into different groups for advertising purposes.

2.3 Reinforcement Learning

Reinforcement learning is a different subset of machine learning where the learning system (agent) can perform different actions and receives rewards or penalties in return. By doing this repeatedly the agent can learn the correct policy, which dictates which action the agent should take in a given situation, to get the most rewards over time.

3 Lectures

Lectures will begin with standard machine learning topics before delving into deep learning. We cover not only classical machine learning and deep learning algorithms, but also new and exciting advances in the field. The club will be split into a beginner series for those new to machine learning and an advanced series for those who already know the fundamentals. The advanced series will be different than it was last year in that it'll focus much more on Kaggle competitions and research instead of lectures.

The following lecture schedule is the current plan for the school year up till Winter Break for the beginner's group. We plan to give discuss major topics, such as Decision Trees and Neural Networks, over the course of multiple weeks. During each two week segment, we plan to give one coding competition and a problem set for rankings (see Sections 4 and 5). Some topics will be easier to understand than others depending on the math involved, but we'll do our best to make lectures understandable while not sacrificing lecture quality. The lecture schedule is subject to change, and the most up to date version can always be found at tjmachinelearning.com/schedule/2021.

4 Problem Sets

Problem Sets will be given occasionally to ensure students learn and retain material. These, along with the In-House contests (see Section 5), will be used to rank students.

5 Competitions

Machine Learning Club will be holding weekly in-house contests through Kaggle Classroom. Students will be ranked based on their achievement in these contests.

For all competitions, we will be using Python 3. If you're unfamiliar with Python or programming, don't worry because it is fairly simple to learn. We recommend starting here: <https://www.python.org/about/gettingstarted/>.

As the year progresses, Machine Learning Club members can participate in real-world Kaggle competitions (kaggle.com/competitions). Substantial prize money is awarded to winners of contests, however, students will be competing against anyone in the entire world, so the probability of winning is extremely low. Nevertheless, Kaggle competitions are a valuable learning experience.

6 The Website

Most information is conveyed through the official Machine Learning Club website, <https://tjmachinelearning.com/>. Here, you can find the lectures along with any presentations, notes, rankings, or additional resources.