# Middle/High School Machine Learning Lab Ideation

General Ideation

[Middle/High School Machine Learning Lab Ideation](#_ra17czxv1xp7)

[Plan](#_s9zxz9hgha7f)

* Each lab needs:
  + Flash/ interesting aspect for a middle school student
  + Specific element of machine learning being taught or stressed

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Machine Learning Lab Ideas** | | | | | | | |
|  | Games | Sports | Art, Music | Lifestyle | Literature | Medicine | Environment/ Society |
| Ideas 1: | ~~The Ai opponent continuously getting better at beating the user in a game such as soccer, FIFA like.~~  **Similar idea can be used in other games (Doctor who mini game, demo tomorrow)**  **Soccer physics type game (also demo tomorrow)** | | Students rate songs from a large database of music and an algorithm will generate a song based on the rankings.  -Applicable to poems or books as well.  -Can be applicable to images and visual effects as well | **Youtube title quality detection + statistics** | Which non stop-words (i.e. “the,” “a,” etc.) appear most frequently in a literary work (this is just basic data analysis)  Learning differences between frequency of word usage across time periods  Recurrent neural network / Lstm | Catering a treatment regime to an individual. For example, typing patterns is an indication of the degree to which Parkinsons is affecting an individual. So if you can measure how effective the medication is, you can modify the dosage on the fly. | **Evolution of the world -- using a variety of factors students can manipulate the outcome earth.**  Cosmos-esk 40 seconds from life to humans  <https://www.youtube.com/watch?v=5g2crxb-PJs>  Variation auto encoder |
| Ideas 2: | Tamagotchi like game with character becoming the ultimate friend | Do something cool with Olympics and statistics? | **Completing an image in some cool way, or completing a song in some cool way.**  <https://deepdreamgenerator.com/feed>  -- Stephen emailed dude | **Alien learning how to be human (stranger things esk) either through literature or watching humans or interaction with users.** | | Improved diagnostics - ML algorithms can take much more data into account, such as symptoms, place of home/work, recent outbreaks of diseases, genetic factors, etc. to give a diagnostic. | Something fun to do with volcanoes. |
| Ideas 3:  Students doing some kind of social good. (high potential for something cool, just don’t know what)  Education: Adaptive tests/assessments to improve learning.  Language: Real-time translation  Environment: Adaptive/Sustainable agriculture  Law: ML to identify human traffickers  Accessibility: Microsoft Seeing AI  **Homelessness:** ML to optimize the distribution of homeless services | | | | | | | |
| Ideas 4: |  | **Fantasy Sports,** | Using ML to write a poem. |  |  |  |  |
| Ideas 5: |  |  |  |  |  |  |  |
| Ideas 6: |  |  |  |  |  |  |  |

* 1:30 meeting topics:
  + Viability of the creation of these labs
  + Discuss questions amal had
  + Discussion about the larger elements in ML, so I can be of some use in the creation of the labs
    - Stats - Prediction
    - Concepts
    - Linear regression
    - Neural network
  + Specific elements of ML that we want to teach
  + What will it take to create these labs, visuals, sounds, images, databases, how can we store that information so students in schools can use it?
    - Conversation for erik probably
  + Integrating Kosbie’s animation framework in a week
  + Finalizing specific ideas/topics we think would be interesting
  + Creating the survey sheet
  + Next steps:
    - Creating a mockup of some labs (visual and code wise)
* Goal of the curriculum:
  + Get students inspired about the power of big data analysis / ML
  + Give them an intuition over basic ML algorithms
  + Give students experience over basic coding of ML and the feeling that even though ML is so magical, they can make it happen!!!

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Intuition | What We Will Teach | Scenario(s) | How Many Labs |
| Decision Stump | * Start with 1D * 2D points, each point has a color, separate * Top-down picture of table, have two classes of objects on it * Can start with a number line | * Put your separator at a particular point (0, or the middle, or something), evaluate how good of a separator it is (a simple way is to look at the fraction of points on each side that is each label, we could potentially provide this function as a black-box helper), and then move 1 unit in the appropriate direction | * Use a sport (like basketball), classify player’s position (center or guard) * Stranger Things: In between vs Real World, average color on page * Pick a player esk. |  |
| Decision Tree (Put a bunch of decision stumps together to make a decision tree) |  |  |  |  |
| Logistic Regression |  |  |  |  |
| Neural Net (many logistic regressions put together) |  |  |  |  |

Background we want students to have:

* Math:
  + Fractions
  + Realm numbers
  + Percentages
  + Axis of 1D and 2D graphs
  + Lines as functions and how to rotate/translate them (students would learn this in geometry, precalc, maybe algebra)
  + Familiarity with linear inequalities
* Computer Science:
  + 1D List
  + For/while loops
  + Conditionals
  + User-defined functions with parameters
* Abstract thinking:
  + Intuition over breaking something down step by step and carrying instructions out in a very mechanical way
  + The notion of variables

|  |  |
| --- | --- |
| Lab Name | Lab Description |
| **Decision Stump 1** | 1D Continuous Decision Stump (with binary labels), example could be sports, students will learn and program the high-level algorithmic idea of: 1) start with some linear separator, 2) evaluate how good it is (maybe we provide this or provide parts of this as a helper), 3) improve upon your linear separator, 4) Repeat from 2 until we have a perfect linear separator |
| **Decision Stump 2** | First Half: 1D Decision Stumps when data is not perfectly separable (big new idea is that you have to keep running the while loop and keep accumulators w=for what the best linear separator is so far)  Second Half: Somehow introduce 2D decision Stumps |
| **Decision Stump 3** | 2D Continuous Decision Stump (with binary labels), we will make it so that the data is linearly separable with a line that goes through the origin (therefore, only rotation/slope matters), we will provide a function rotateUp / rotateDown and then its the same computational thinking as above (initialize with a horizontal line). |
| **Decision Tree 1** | 1D decision trees |
| **Logistic Regression** |  |
| **Neural Net** | Compression, (better idea to relate to kids, auto encoder) |

Knn

Linear Regression

Convolutional neural network-

Generative adversarial network

~1:15 labs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Labs | Theme | Paths | Activity | Learning Goals: |
| 1.1 | Classification | Decision Stump (1D linearly separable, 1D non-linearly separable) | * Change the example   + <https://docs.google.com/presentation/d/1r2O7IdFXa8dKT8Iw37v5w2LMzTMIoq6luQGSpWoY49s/edit?usp=sharing> * Disease detection/crime detection coding activity | * The algorithmic idea of: 1) try a separator, 2) evaluate how good it is, 3) improve it * Pure vs. Impure Classification (even if you can’t learn a perfect machine learning algorithm, you can learn one that is good enough) |
| 1.2 | Classification | Decision Trees (2D) | * 1 part lecture teaching ideas * Disease detection more advanced | * How to pick your battles: tackling the “biggest gain” features first. * Chaining multiple machine learning algorithms together to make a more powerful algorithm. |
| 1.3 | Classification | Application of Decision Trees | * Long (open-ended ish) coding activity * Potentially moving into higher dimensional data |  |
| 2.1 | Classification K-Nearest Neighbors | K Nearest Neighbors | * Lecture * Talk about how different approaches affect effectiveness of model. * Coding Activity (possibly evolution) * Crime analysis/ geography example | * There are multiple machine learning approaches to similar types of problems -- it is crucial to use knowledge of the application domain as well as knowledge of the algorithm to determine what is best. * Introduce the notion of overfitting? KNN really overfits to the data, whereas decision trees don’t |
| 2.2 | Classification K-Nearest Neighbors | Crime/Disease Detection Part 1 | * Based off of [Dr. John Snow’s analysis](https://www1.udel.edu/johnmack/frec682/cholera/) that revealed that cholera is caused by water |  |
| 2.3 | Classification K-Nearest Neighbors | Crime/Disease Detection Part 2 | * Based off of [Dr. John Snow’s analysis](https://www1.udel.edu/johnmack/frec682/cholera/) that revealed that cholera is caused by water |  |
| 3.1 | Weighted Majority | Prediction with Expert Advice | * Friends (experts) are suggestions weather to watch a movie or not | * The notion of upweighting and downweighting objects/data based on how preferable they are |
| 3.2 | Weighted Majority | Prediction with Expert Advice | * You have multiple classifiers (Decision Trees and KNN) and are trying to determine which are best are predicting malaria | * You can take multiple machine learning algorithms and combine them into a more powerful ML algorithm (this is not a cascading combination like decision trees, but rather a learning which algorithms to trust) |
| 4.1 | Neural Nets | Introduction & Cool Applications | * General Lecture * How does everything we’ve learned build into it? * Start training in the beginning of the class and by end have it working. | * There is so much more cool stuff to learn in ML and cool applications!!!! |
| 5.1 |  | Projects |  |  |
| Projects for remainder of days... | | | |  |

~~TODO:~~

1. ~~Stephen: Start on Classification, see how far you get~~
2. ~~Amal: Start on KNN technical part, see how far you get~~
3. ~~Chris: start of KNN activity part, see how far you get~~

Java to Python

For loops

Binary search

Dictionary Maps

Lists

* 1.1, 2.1, 3.1
  + Longer lectures to introduce topics in depth
  + One off coding activities specified to the learning material
* 1.2, 1.3, 2.2, 2.3, 3.2
  + Shorter lectures to reinforce topics
  + Coding activities following theme
* 4.1
  + Lecture on interesting aspects of neural nets and how

Jiggle ziggle

Amal

Stephen

Chris

# Plan

|  |  |  |  |
| --- | --- | --- | --- |
| Lab | ML Topic | Lecture | Coding Activity |
| 0 | Java->python |  | Practice Changing code i.e. binary search, etc. |
| 1.1 | Decision Stump | [Slide](https://docs.google.com/presentation/d/1r2O7IdFXa8dKT8Iw37v5w2LMzTMIoq6luQGSpWoY49s/edit?usp=sharing) | Disease detection |
| 1.2 | Decision Trees (2D) | **Transition: what if your data is not linearly separable?** Short Powerpoint | Writing general purpose decision tree (scaffolded) |
| 1.3 | Application of Decision Trees |  | Disease detection |
| 2.1 | Weighted Majority | Transition: Offline vs. online (taking data. Take multiple aspects of data and combine. Longer Powerpoint | Deciding whether to go to a movie/concert/etc. based on friends’ recommendations (weights are how much to trust a friend)  [Sample Solution](https://drive.google.com/a/andrew.cmu.edu/file/d/1iqB-Jyf65RvZKYqM_ntWf8aj8_M6u7EP/view?usp=sharing) |
| 2.2 | Weighted Majority | Short Powerpoint  Mostly a recap. | Generalizing the WMA, reading data from a file and visualizing how good the algorithm is.  [Sample Solution](https://drive.google.com/a/andrew.cmu.edu/file/d/1A7LEokXDyco4aUqpMbe4S3BeVtJtPncL/view?usp=sharing) |
| 2.3 | Weighted Majority + Decision Trees | Maybe a powerpoint? | Combining Decision trees and wma to do disease detection -- the experts are different classifiers. Figure |
| 3.1 | K Nearest Neighbors | Longer Powerpoint | KNN Algorithm and code (Theme TBD) |
| 3.2 | Weighted Majority + KNN | Transition: 2D -> just distance. **Do we incorporate weighting?** Short Powerpoint recapping weighted majority + Knn. Coolmuffins. | Predicting survival based on DNA |
| 4.1 | Neural Nets | Longer Powerpoint. Have them actually emulate a neural network. | Show them a neural net (online). **An activity where they are a neural net? (probably not coding)** |
|  | **Hackathon** |  |  |
|  |  |  |  |

* Everyone:
  + What we need with theme:
    - 5 Coding activities to fit into learning goals
    - Examples for within lectures
* Amal: also start writing Weighted Majority Algorithm Code
* Stephen: Also continue with slides, and change the table to a number line

Current Theme:

Malaria

Detective (maybe)

Themes Ideas:

Superheros- make their own superhero?

Video games

Camping-

Solar System-

Sports-

Future-

Past-

Movies-

Time Travel-

Self driving cars-

Friends-

Online dating-

Karate-

Cooking-

Puzzles?