Introduction

Week	TUES. Week 1
Assignment Due	
Assignment Done	☐
Due Date	
Notes Done	☑

Presentation

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/55d35157-ed8d-4445-93d5-2e3ec6838310/L1-Intro.pdf

Class Notes

What is ML?

Machine learning - the study of computer algorithms that improve automatically through experience

- Subtopic in AI
- Formally:
  - Improve at task T
  - With respect to performance measure P
  - Based on experience E
- Ex: Recognize a person
  - T: recognize a person

- P: # of time we recognized a person correctly
- E: database of labeled races

AI deals w/ any “intelligent” task performed by a non-human agent → often “path finding algorithms”

- ML specifically deals with making decisions based on acquired data, both past and current

How Do We Do It?

Need to think about:

1. What needs to be learned?
    a. Drives us to think about what algorithms we need to use
2. What feedback can we get in what form?
    a. Supervised learning (correct answers for each example)
    b. Unsupervised learning (correct answers NOT given)
3. What representation should we use (features)?

Problems in ML

There is no single algorithm currently that rules them all:

- Some things can drive us to certain algorithms such as if data is labeled, amount of data, nature of data, but as of now, there are no definitive rules on what is the best algorithm.
- Different tasks, most common:
    - **Clustering** → having raw data, not sure what it means, trying to find patterns within the data to group by similarity measurements
    - **Regression** → want to compute values
        - Based on a bunch of info, can we compute an outcome value?
    - **Classification** → given data, can we predict what the data is? what class does it pertain to?

Administrative Points

- 5 Homework Sets, includes theoretical questions as well as implementation of ML algos
  - Late policy: 1% off for every hour late, max of 48 hours
    - Can use 1 late pass in course, prior to ONE assignment's due date
- Midterm/Final
  - What theoretical parts of HW is preparing us for

Homework Assignment

We'll have the material/knowledge that we need to complete this assignment by next Tuesday's lecture completion.

All necessary Python packages are installed accordingly on Tux!

Fundamental ML Concepts

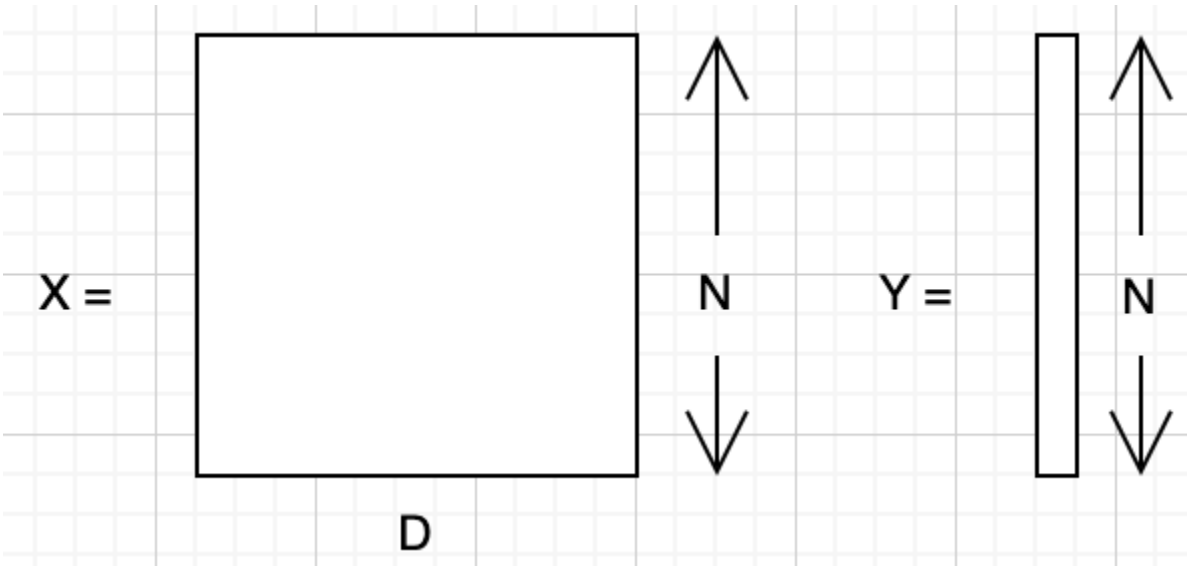
ML Overview

We can break ML tasks into two categories:

- **Supervised learning** - Data X_i and correct answer (label) Y_i given for each example $i \in \{1, ..., N\}$
  - Clustering
  - Regression
- **Unsupervised learning** - Only data given for each example, X_i
  - Clustering

Observable data is typically stored in a matrix X with n rows and d columns.

- Each column is a feature, which is a piece of information about the observation.



- X_1 = first row of data

Types of Data

Features can fall into one of three categories:

- ***Categorical-Nominal (unordered)***
  - Ex: Car model, school
- ***Categorical-Ordinal (can be ordered)***
  - Is there a logical order to a categories provided?
  - Ex: Colors, small < medium < large
- ***Continuous Valued***
  - Ex: Blood pressure, height

CS383 ML Algorithms

Next Tuesday, we'll be learning about feature reduction (feature selection, feature projection).

- At times, having more info about each observation can make things difficult. We may want to reduce the amount of info that we have in an intelligent way to not discard useful information.

Then, we'll go into classification, going into simpler algorithms:

- There’s a natural transition to feature reduction algorithms to linear discriminant analysis (LDA) and nearest neighbors algorithms (KNN).
- Then we’ll enter medium-difficulty algorithms, such as decision trees (DTs), statistical classification algorithms, and logistic regression.
    - We’ll then be lead into an introduction to artificial neural networks (ANN) → we dive into this into deep learning course
    - We’ll then learn about support vector machines (SVMs) and hidden Markov models (HMMs).

We’ll enter linear regression, then clustering algorithms such as:

- K-means
- Agglomerative clustering
- Mixture models