Experimental Design for Machine Learning on Multimedia Data
Lecture 2

Dr. Gerald Friedland,
fractor@eecs.berkeley.edu

Logistics

- Office Hours:
Gerald Friedland
Monday 1pm-2pm 424 SDH
Rishi Puri: TBD

- Class is now a 3 unit class

- It counts as technical elective for MEng.

- We will have a final exam for MEng.

- BCourses and Piazza is setup

Project Requirements

- See webpage:
  http://www.icsi.berkeley.edu/~fractor/spring2019/

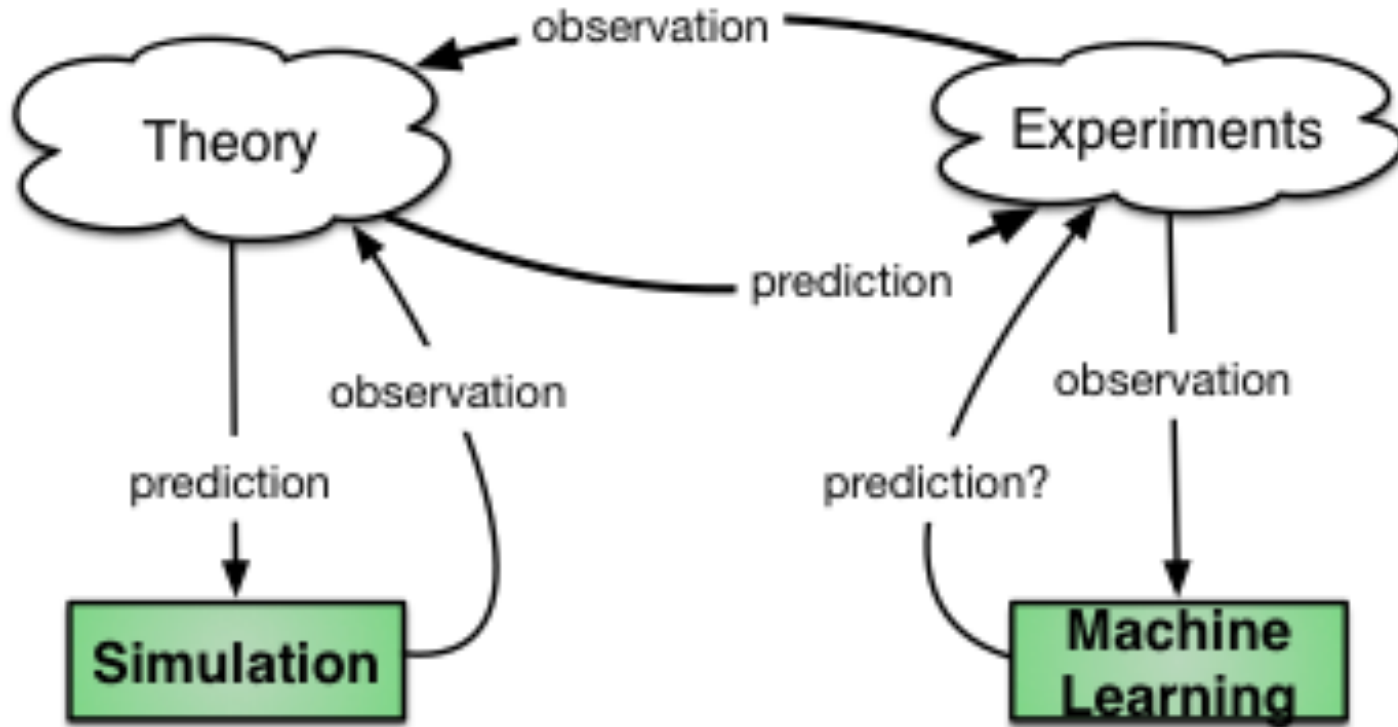- Project is due final day of RRR week. Need final week for grading.

Today

- Repeat: The scientific process and machine learning
- A quick trip into epistemology
- Reminder: Mathematical Fundamentals
- A quick trip into statistics: Galton Board
- Question 1: Statistic or deterministic machine learner? When to prefer which.

The Scientific Method

Data Science: The Science of Automating the Scientific Method

Epistemology

the theory of knowledge, especially with regard to its methods, validity, and scope. Epistemology is the investigation of what distinguishes justified belief from opinion.

Goal of Science

The goal of science is to predict observations given previous observations.

Example: Newton's Apple (see whiteboard)

We can only predict repetition (redundancy)!

Occam's Razor

- Among competing hypotheses, the one with the fewest assumptions should be selected.

For each accepted explanation of a phenomenon, there may be an extremely large, perhaps even incomprehensible, number of possible and more complex alternatives, because one can always burden failing explanations with ad hoc hypotheses to prevent them from being falsified; therefore, simpler theories are preferable to more complex ones because they are more testable.

(Wikipedia, Sep. 2017)

Important Concepts

- Hypothesis
- Theory
- Theorem
- Lemma
- Proof
- Observation
- Axiom
- Assumption
- True and False
- Contradiction

Reasoning: Logic

- First order logic, second order logic: Short descriptions of a decision tree.

- Bayesian Inference: Reasoning using only the topology of the decision tree.

- See whiteboard

Math basics for this class

- Information Content
- Entropy (Hartley, Shannon)
- Bit-arithmetic
- Deterministic Automata
- Pigeonhole Principle
- Data Processing Inequality

See whiteboard

Statistics

- See Galton Board

Which type of Machine Learner to chose

- Chose a statistical machine learner for predictions where you have too little information. (e.g. betting game)

- Chose a deterministic machine learner for predictions where you have too much information. (e.g. image classifier)