CSE 417T
Introduction to Machine Learning

Instructor: Chien-Ju Ho

Plan for today

- Welcome and introduction
- What's the class about?
- Logistics

- Lecture
  - Setting up the learning problem
  - Perceptron learning algorithm

- Homework 1 will be announced before early next week

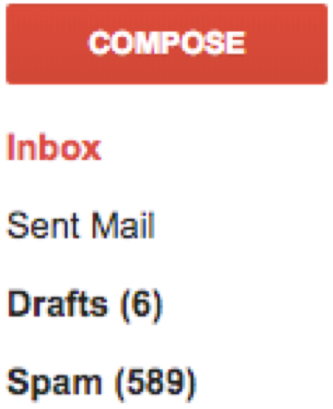
About Me

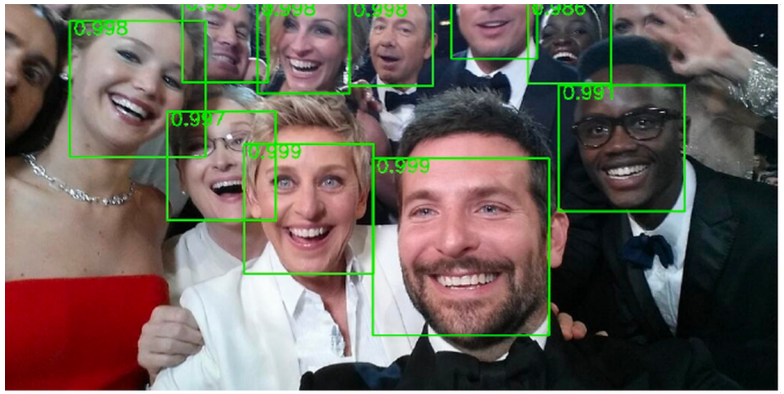
- Joined WashU in Fall 2017.

- Research interests:
  - Design and analysis of human-in-the-loop systems
  - Crowdsourcing and human computation, machine learning, game theory, optimization, online behavioral social science, and human-computer interactions.

What is Machine Learning?

What Machine Learning Can Do?





A Simpler Example: Credit Card Approval

age	32 years
gender	male
salary	40,000
debt	26,000
years in job	1 year
years at home	3 years
...	...

Should the bank approve the credit card application?

- A brute-force solution
  - Build a large table that maps all possible attribute combinations to a credit approval decision
- Not really feasible…
  - Possible attribute combinations could be infinite
  - Storage, computation, how to come up with the table?

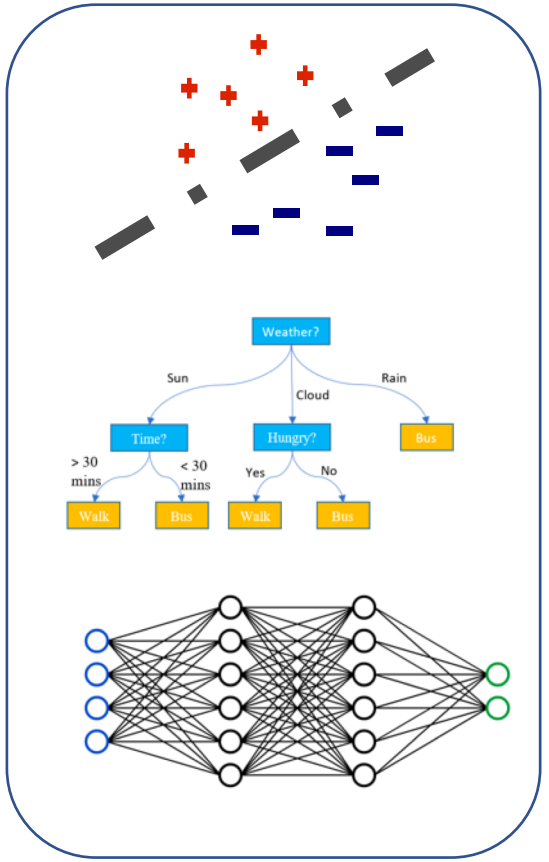
A ”Machine Learning” Approach

Banks have lots of data.

- customer information: salary, debt, etc.
- whether or not they defaulted on their credit.

Data (Historical Customer Data)

Hypothesis / Model (Some math function)



Find a hypothesis that “fits” the data
(The process requires a lot of computation)

What is machine learning?

"using a set of observations to uncover an underlying pattern"

"learning from data"

Use scenarios of machine learning

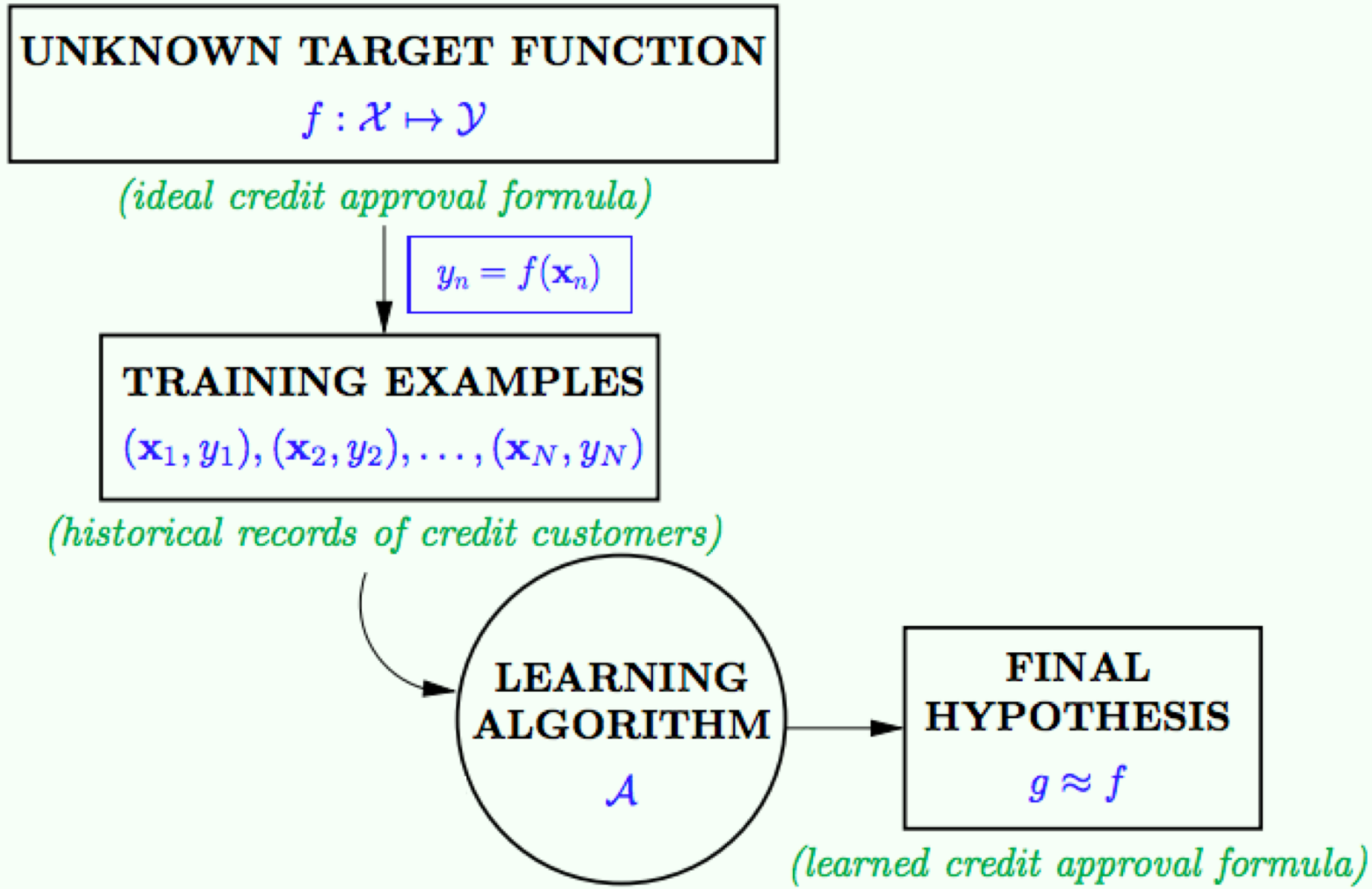
- A pattern exists
- No analytical solution: We cannot pin it down mathematically
- We have data on it

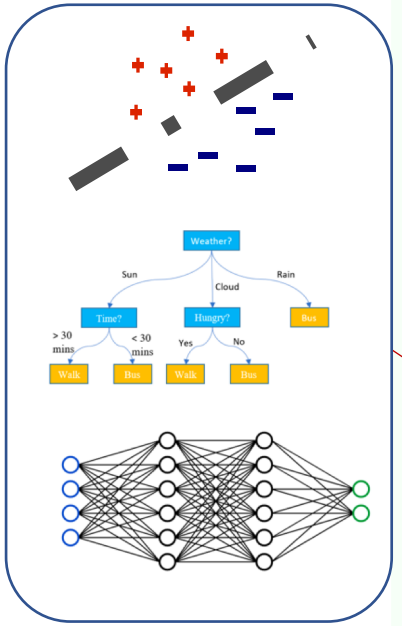
More formally for (supervised) learning

- Formulation: (credit card approval example)
  - input (features): $\boldsymbol{x} \in X$ (customer's information)
  - output (label): $y \in Y$ (good/bad customer)
  - unknown target function: $f: X \to Y$ (ideal credit approval formula)
  - data $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)$ (historical records)

  - goal: learn a g close to f (formula to be used)

- Two central questions
  - How do we learn g?
  - What can we say about how close g is to f?


UNKNOWN TARGET FUNCTION
$f : \mathcal{X} \mapsto \mathcal{Y}$
(ideal credit approval formula)
$y_n = f(\mathbf{x}_n)$
TRAINING EXAMPLES
$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)$
(historical records of credit customers)
LEARNING ALGORITHM
$\mathcal{A}$
FINAL HYPOTHESIS
$g \approx f$
(learned credit approval formula)



UNKNOWN TARGET FUNCTION

$f : \mathcal{X} \mapsto \mathcal{Y}$

(ideal credit approval formula)

$y_n = f(\mathbf{x}_n)$

TRAINING EXAMPLES

$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)$

(historical records of credit customers)

LEARNING ALGORITHM

$\mathcal{A}$

FINAL HYPOTHESIS

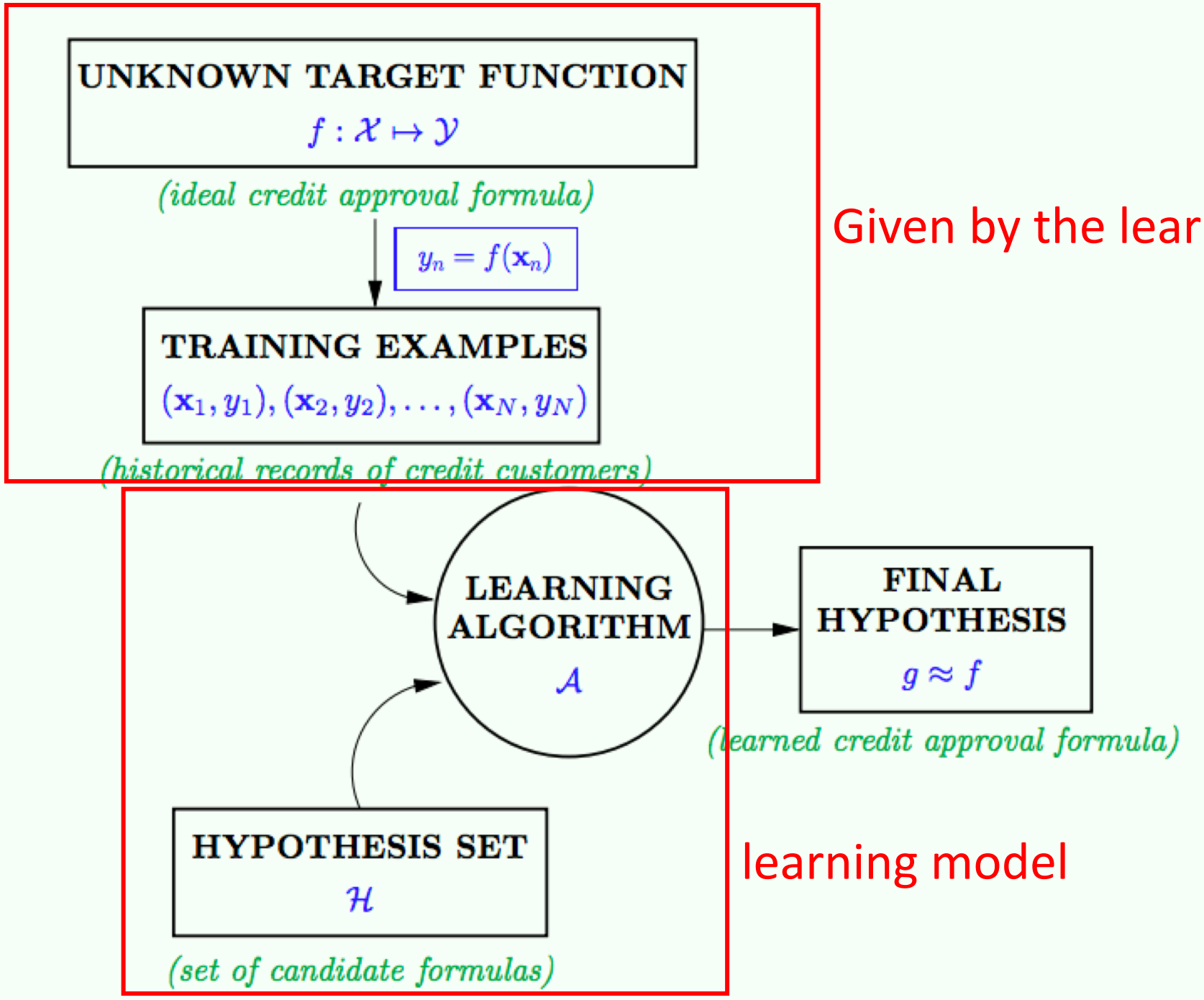
$g \approx f$

(learned credit approval formula)

HYPOTHESIS SET

$\mathcal{H}$

(set of candidate formulas)



Given by the learning problem

learning model

Course Plan

- First half of the semester: **Foundations**
  - Focus on **linear models**
    - fundamental components of many others models
  - Discuss the theoretical foundations of machine learning
    - Heavy use of probability, linear algebra, and optimization
- Second half of the semester: **Techniques**
  - Discuss different learning models

Course Plan

- Foundations
  - What's machine learning
  - Feasibility of learning
  - Generalization
  - Linear models
  - Non-linear transformations
  - Overfitting and how to avoid it
    - Regularization
    - Validation

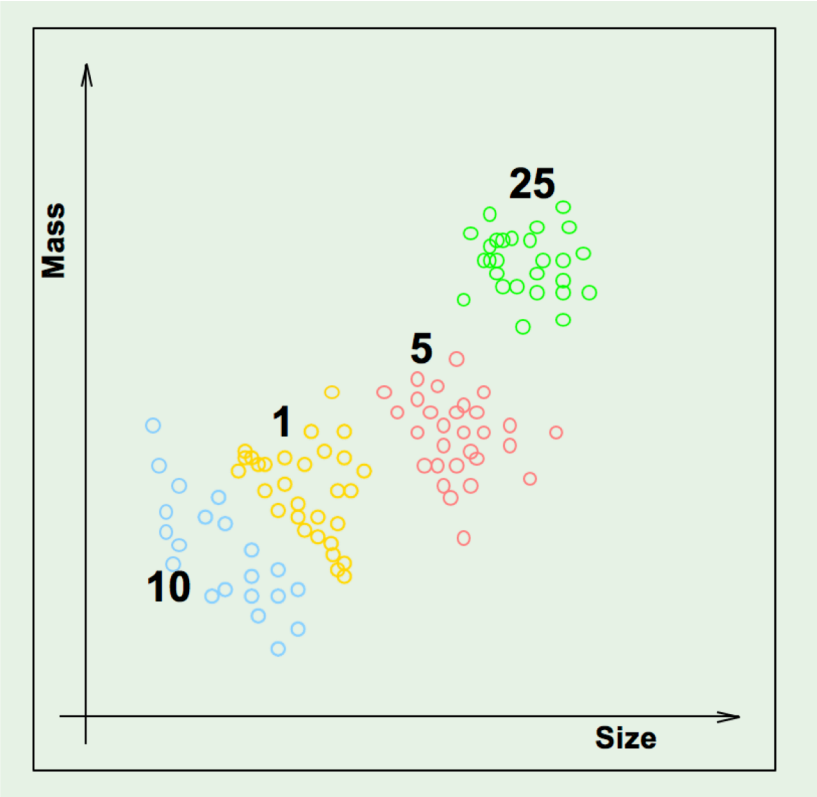
- Techniques
  - Nearest neighbors
  - Decision tree
  - Support vector machine
  - Boosting
  - Random forest
  - Neural networks
  - …

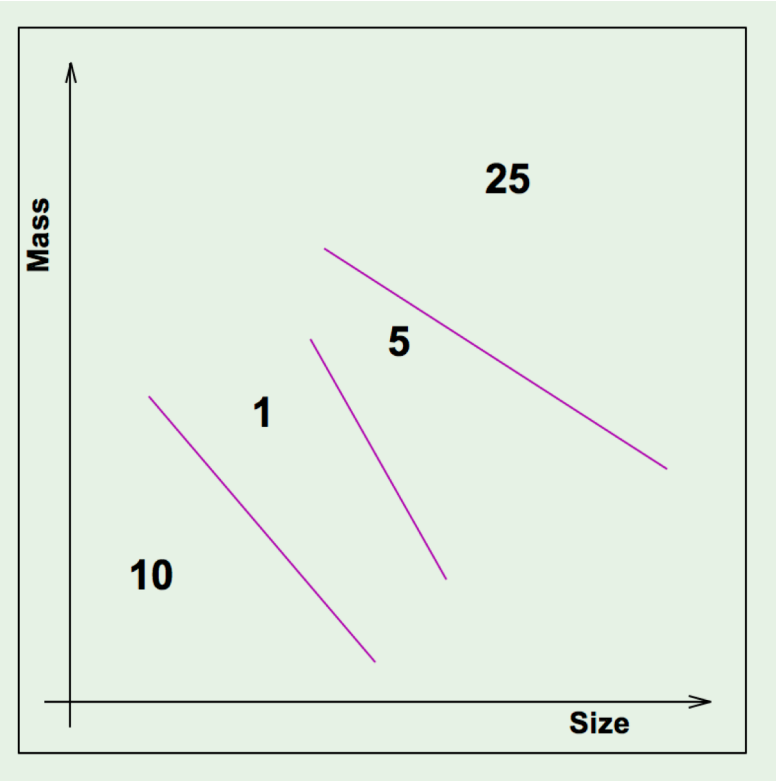
Types of learning

- Supervised learning (main focus of this course)
  - Given training data (input, correct output),
  - try to predict the output for data not seen before
- Unsupervised learning
  - Find patterns in data given only (input)
- Reinforcement learning
  - Learn how to act, based on rewards for actions

Supervised learning

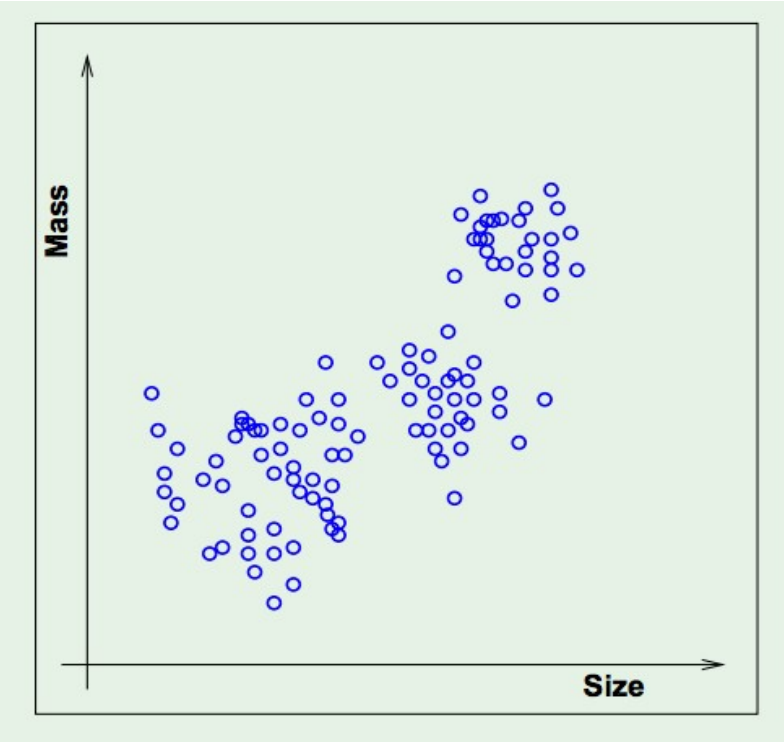
- Given data with (input, correct output), learn a pattern that can predict previously unseen data

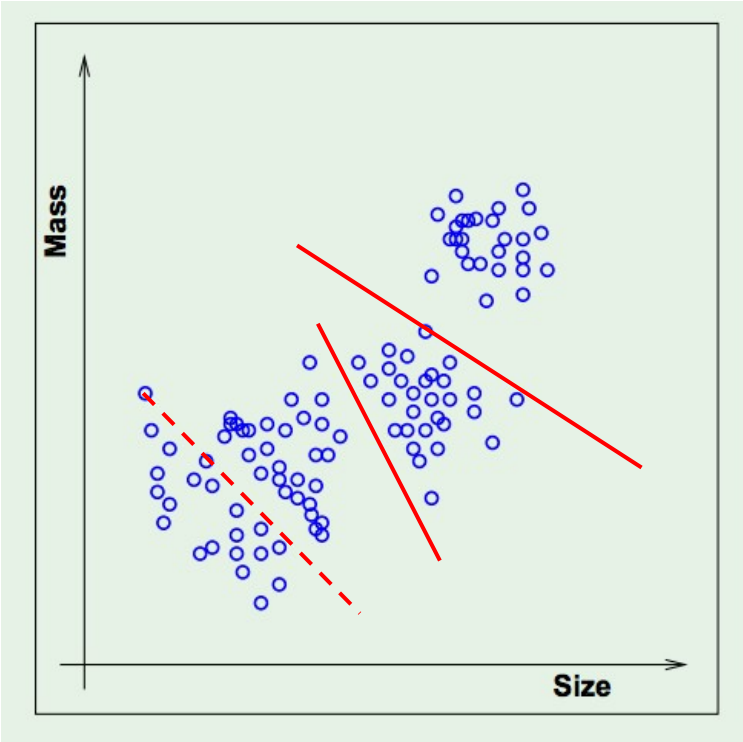




Unsupervised learning (more in 517A)

- Suppose you only have the feature vectors but no labels. Still want to describe the data in some useful way.

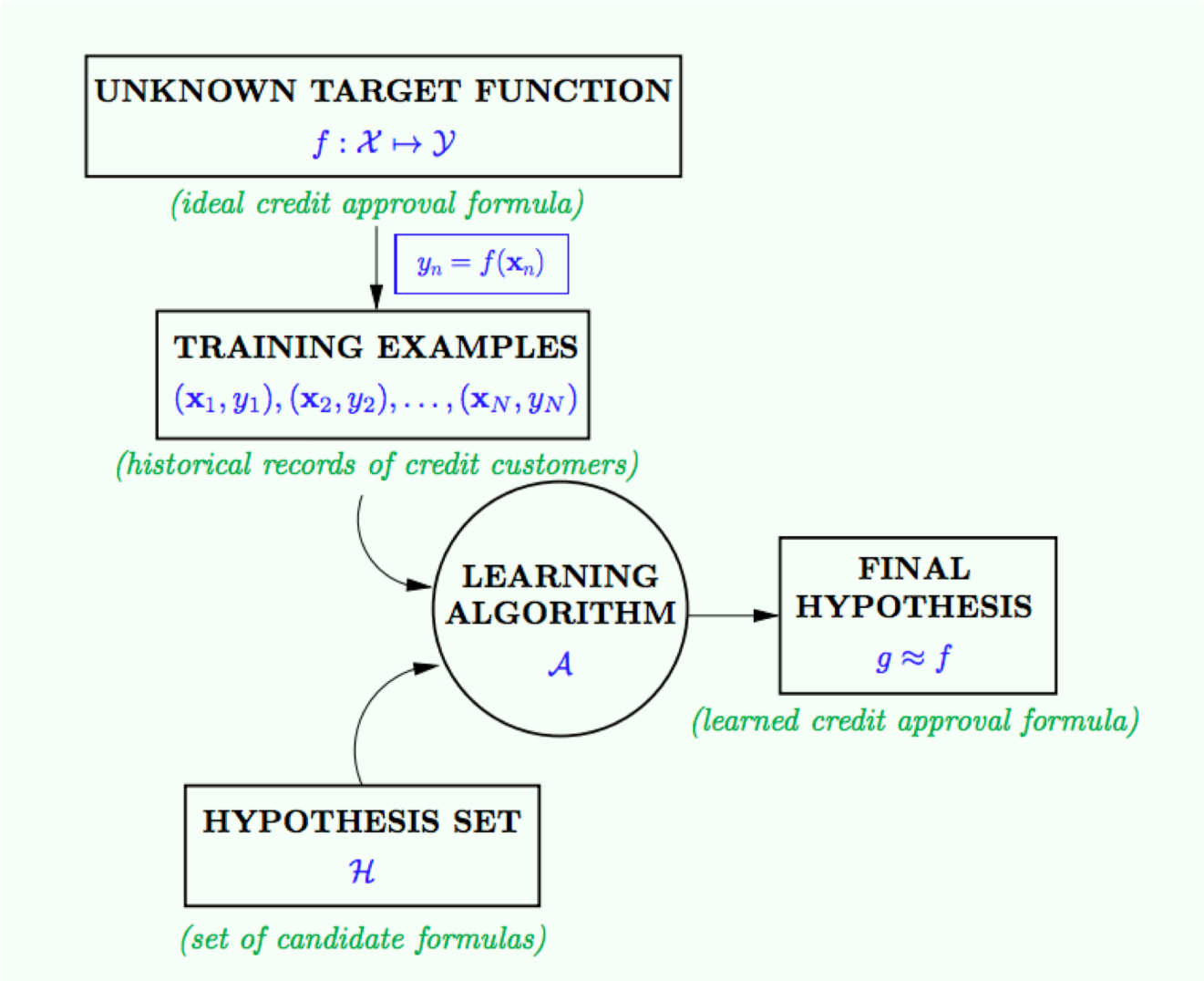




Reinforcement learning (more in 511A)

- Agent interacts with the world by taking actions
- Feedback is in the form of rewards (or costs)
- Agent must learn a policy, which maps from the state of the world to an action
- Major issues:
  - Delayed reward / credit assignment
  - Exploration / exploitation

Course plans (focus on supervised learning)



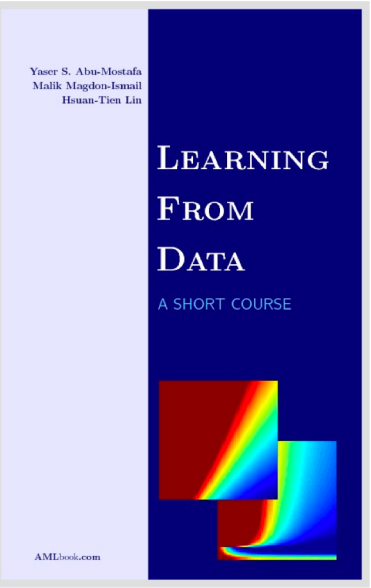
Logistics

- Websites
  - Course website: http://chienjuho.com/courses/cse417t
  - Piazza for discussion.
  - Gradescope for homework submissions.
  - You are responsible for following the announcements on website and Piazza.

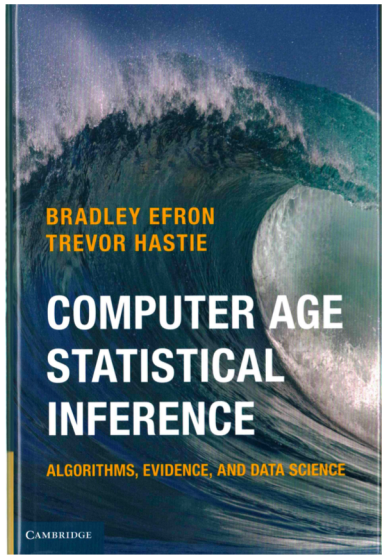
- TA and Office Hours
  - There will be 12 TAs
  - Office hours will be announced in the 2nd week and start in the 3rd week

Course Information: Textbooks

- *Learning From Data*
  - Y. Abu-Mostafa, M. Magdon-Ismail, and H-T Lin.
  - http://amlbook.com/
  - We will go through this book in the first half of the semester.



- *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*.
  - B. Efron and T. Hastie
  - https://web.stanford.edu/~hastie/CASI/
  - We will reference a few sections as the course materials. The PDF file is freely available on their website



Course Information: Grading

- Homework assignments (5 to 7): 50%
  - Mix of programming and pencil-and-paper problems
  - Worst score discounted by 50%
  - 5 total late days, no more than 2 on any one assignment
  - Use **MATLAB** as the programming language

- Two in-class exams: 50% (25% each)
  - Feb 27 or March 3, 2020 (Tentative)
  - April 23, 2020

Course Information

- Let’s look at the syllabus
  - Collaborations
    - You are encouraged to discuss homework with other students
    - **Must write your own solutions**
    - **Must cite all external sources (including other students)**
  - Academic integrity
    - Zero tolerance on the violation
    - Will be reported to the university
    - There will be permanent record if found guilty
  - Other accommodation resources

Course Information

- Questions not covered in the syllabus?
  - Ask me!
  - Generally I don’t grant individual exceptions
    - Can I do extra work to get more points?
    - I have another exam this week, can I get an additional day for the assignment?
    - I work really hard but can’t finish the assignment on time. Can I get an additional day for the assignment?
    - I work really hard. Can I get higher grades?
    - **No** to all the above.
    - Exceptions: family/medical emergencies
  - Rule of thumb:
    - I only say yes if I feel comfortable giving the same treatment to everyone in the class.

Is this course for you?

- This is going to be a very theory-heavy course
  - The “T“ in CSE 417T stands for “Theory”
  - There will be **A LOT OF** math!

- We focus on understanding the foundations of machine learning

- If your main goal is to learn how to apply ML in your problems, this might not be the best course for that.
  - Check out CSE 217A, ESE 419, INFO 577

Waitlist and Enrollment

- If you are on the waitlist and want to get enrolled:
  - Complete homework 0 (posted on the website)
    - Prove PLA (discussed next) converges in finite steps
    - Explain why you want/need to take this course in this semester
  - Submit via Gradescope by 11am next Tuesday (Jan 21)

  - Priorities will be given to students who benefit the most by taking this course this semester (condition on satisfactory answer to the questions)

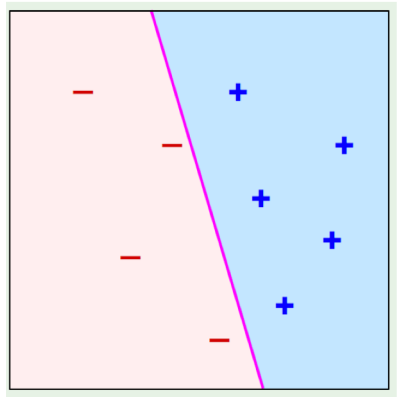
- If you are enrolled
  - You should look at homework 0 as well (please don't submit yet)
  - It helps you get a better sense of what this course is about

Brief lecture notes

Linear hypothesis space (Perceptron)

- Input $\vec{x} = (x_1, x_2, \ldots, x_d)$
- Output $y \in \{-1, +1\}$

- A hypothesis h is characterized by
  - weight vector $\vec{w} = (w_1, \ldots w_d)$
  - threshold b

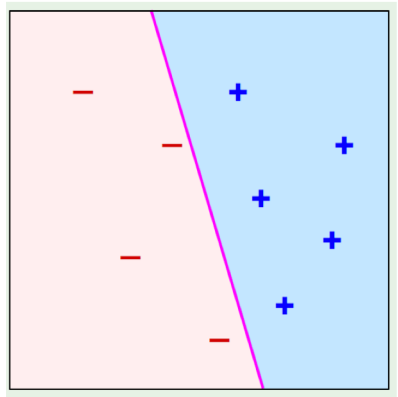


- $h(\vec{x}) = sign\left(\sum_{i=1}^{d} w_i x_i - b\right) = sign(\vec{w}^T \vec{x} - b)$
  - Predict $+1$ if $\vec{w}^T \vec{x} > b$
  - Predict -1 if $\vec{w}^T \vec{x} < b$

Linear hypothesis space (Perceptron)

- To simplify the presentations, define
  - $x_0 = 1$
  - $w_0 = -b$

- And we implicitly let
  - $\vec{x} = (x_0, x_1, \ldots, x_d)$
  - $\vec{w} = (w_0, w_1, \ldots, w_d)$



- A hypothesis can then be written as
  - $h(\vec{x}) = sign(\vec{w}^T \vec{x})$
  - We will interchangeably use h and $\vec{w}$ to express a hypothesis in Perceptron

Perceptron Learning Algorithm (PLA)

- Given a dataset $D = \{(\vec{x}_1, y_1), \dots, (\vec{x}_N, y_N)\}$
- Assume the dataset is **linearly separable**

- Perceptron Learning Algorithm
  - Initialize $\vec{w}(0) = \vec{0}$
  - For $t = 0, \dots$
    - Find a misclassified example $(\vec{x}(t), y(t))$ in D
      - That is, $\text{sign}(\vec{w}(t)^T \vec{x}(t)) \neq y(t)$
    - If no such sample exists
      - Return $\vec{w}(t)$
    - Else
      - $\vec{w}(t+1) \leftarrow \vec{w}(t) + y(t)\vec{x}(t)$

Notation:
We use $\vec{w}(t)$ to denote the value of $\vec{w}$ at step t of the algorithm.

Similarly, we use $(\vec{x}(t), y(t))$ to denote the data point found at step t.

Perceptron Learning Algorithm (PLA)

- Theorem (informal):
  - If D is linearly separable, PLA find a linear separator that separates the data in D within a finite number of steps.

- You will need to prove this in the homework