ISTA 421 + INFO 521
Introduction to Machine Learning

Lecture 1: Introduction

Clay Morrison
claytonm@email.arizona.edu
Harvill 437A
Phone 621-6609

21 August 2017

1

Welcome to ISTA 421 + INFO 521

- Today:
  - Introductions
  - Syllabus
  - Course Structure & Goals
  - Intro to ML

Course Website (also linked from the D2L course webpage)
http://w3.sista.arizona.edu/~clayton/courses/ml

2

Your Instructors

- **Clay Morrison**
  - Harvill 437A, 621-6609
  - claytonm@email.arizona.edu
  - Office Hours:
    - 2-3pm Wednesday
    - By appointment (email me!)

- **TA: Enrique Noriega**
  - Harvill 456
  - enoriega@email.arizona.edu
  - Office Hours:
    - 4-5pm Monday, Wednesday

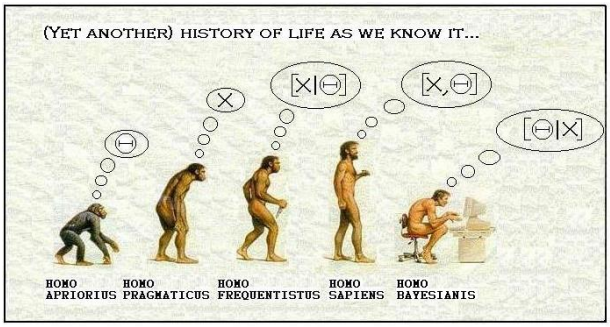
Please contact us ahead of time if you plan to attend office hours!

General Email Rule: start subject line with [421] or [521] (doesn't matter which)

3

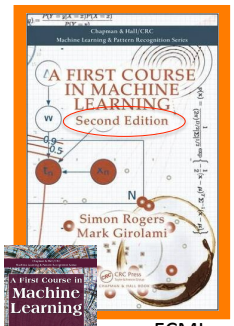
Course Goals

- Basic literacy in core, modern ML methods
- Practical experience implementing ML algorithms and using them on data

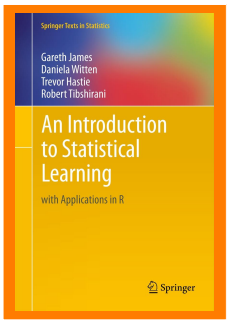


4

Course Texts

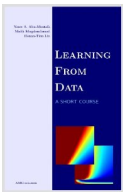


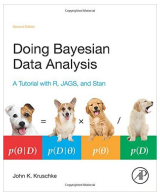
FCML

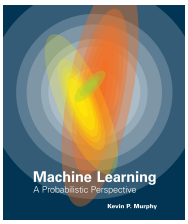


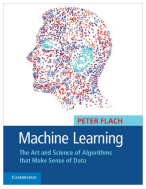
ISLR

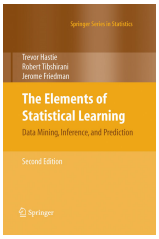
Tom Mitchell, 1997

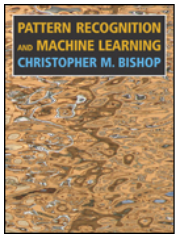


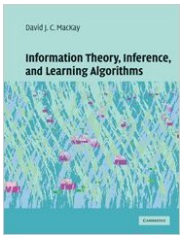


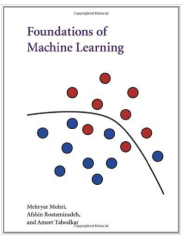


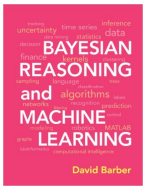






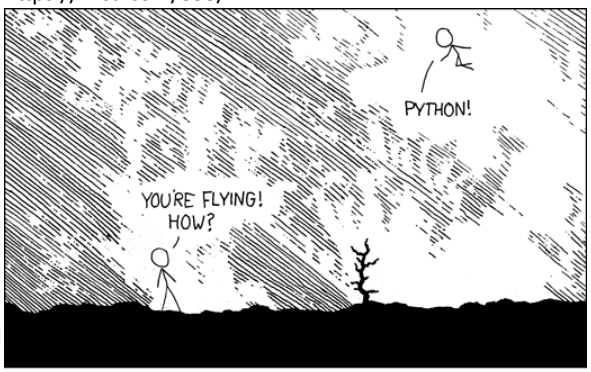






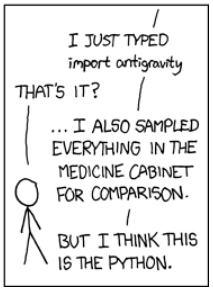
5

https://xkcd.com/353/



Python 2.7 or 3.6
numpy
scipy
matplotlib

All provided assignment code in the course will be in Python 3.6; recommend using 3.6!



http://w3.sista.arizona.edu/~clayton/courses/ml/python_setup.html

6

Topics

- The linear model
  - Regression, Classification
- Classification
  - Probabilistic:
    - Bayes Classifier, Naïve Bayes
    - Logistic Regression
  - Other, *non*-probabilistic
    - K-nearest neighbors
    - Support Vector Machines and kernel methods
- Clustering
  - K-means
  - Mixture Models and EM
- Other Unsupervised methods:
  - Principle Components Analysis
  - Latent Variable Models
- Additional topics (time permitting)
  - Neural networks, Deep networks
  - Nonparametric Bayesian Models
    - Gaussian processes, topic modeling
- Probability
  - Quantifying uncertainty
  - Bayesian Approach: Prior, Marginal Likelihood, MAP
- Inference Methods
  - Least Squares
  - Maximum Likelihood
  - Bayesian Inference: Direct and Sampling
- Machine Learning algorithm evaluation
- Learning theory
- Feature Selection and Model Selection

7

What is Machine Learning?

- The goal of machine learning is to build computer systems that can adapt and learn from their experience. (Dietterich, 1999)
- Machine learning usually refers to changes in systems that perform tasks associated with artificial intelligence. Such tasks involve recognition, diagnosis, planning, robot control, prediction, etc. (Nilsson, 1996)
- Some reasons for adaptation:
  - Some tasks can be hard to define except via examples
  - Adaptation can improve a human-built system, or track changes over time
- Goals can be autonomous machine performance, or enabling humans to learn from and understand data (data mining and modeling)

8

Ack: this and some following content adapted from Chris Williams 2006

Some of the Roots of Machine Learning

- **Philosophy**: epistemology, philosophy of science, logical inference: *the Problem of Induction*
- **Mathematics**
- **Physics**: statistical mechanics
- **Statistics**: statistical inference, frequentist & Bayesian
- **Psychological** models (of learning and development)
- **Brain** models, e.g. neural networks
- **Artificial Intelligence**: e.g., discovering rules using decision trees, inductive logic programming, autonomy
- **Engineering**: Statistical pattern recognition, operations research, adaptive control theory

9

From Rob Tibshirani

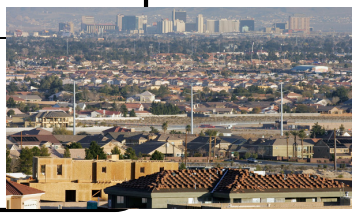
What is the difference between Machine Learning and Statistics?

Glossary

Machine learning	Statistics
network, graphs	model
weights	parameters
learning	fitting
generalization	test set performance
supervised learning	regression/classification
unsupervised learning	density estimation, clustering
grant = $1,000,000	large grant=
place to have a meeting:	nice place to
bird, Utah, French Alps	Las Vegas in

Greater focus on
- prediction
- analysis of learning algorithms (not just large dataset issues)

Greater focus on
- understanding data in terms of models
- interpretability, hypothesis testing



Typical Machine Learning Workflow

(0) **Define the problem.** What is the task we want to teach a computer to do?

(1) **Collect data.** Gather data for training and testing sets. The larger and more diverse the data the better.

(2) **Design features.** What kind of features best describes the data?

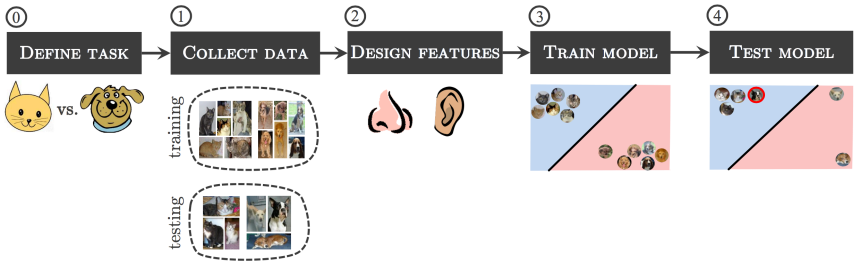
(3) **Train the model.** Tune the parameters of an appropriate model on the training data using numerical optimization.

(4) **Test the model.** Evaluate the performance of the trained model on the testing data. If the results of this evaluation are poor, re-think the particular features used and gather more data if possible.

From : Watt, Borhani & Katsaggelos, *Machine Learning Refined*
http://mlrefined.wix.com/home-page

11

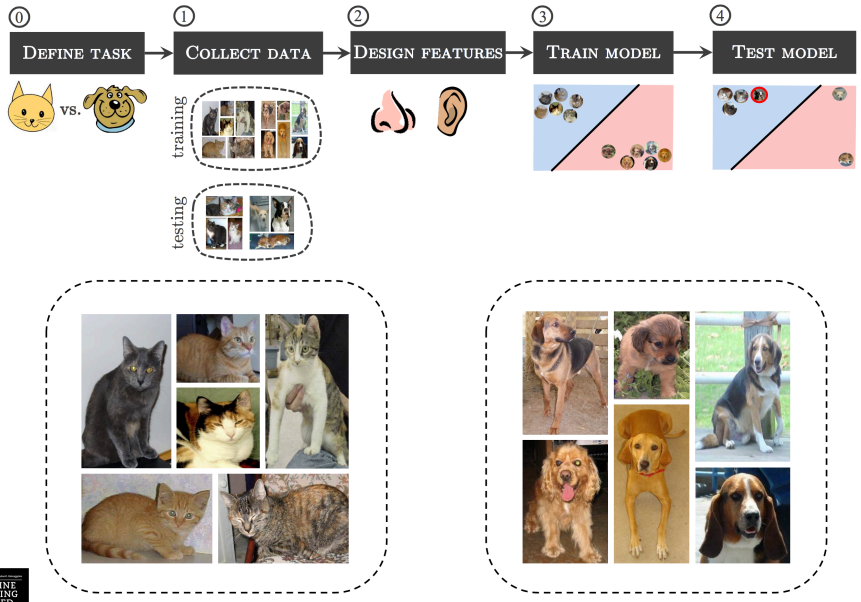
Typical Machine Learning Workflow



From : Watt, Borhani & Katsaggelos, *Machine Learning Refined*
http://mlrefined.wix.com/home-page

12

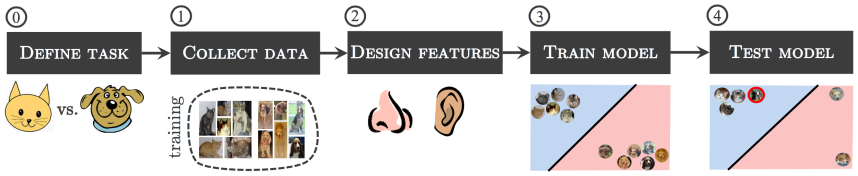
Typical Machine Learning Workflow

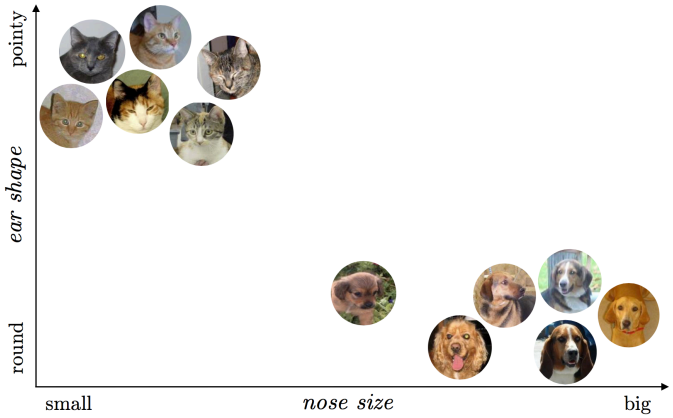


From : Watt, Borhani & Katsaggelos, *Machine Learning Refined*
http://mlrefined.wix.com/home-page

13

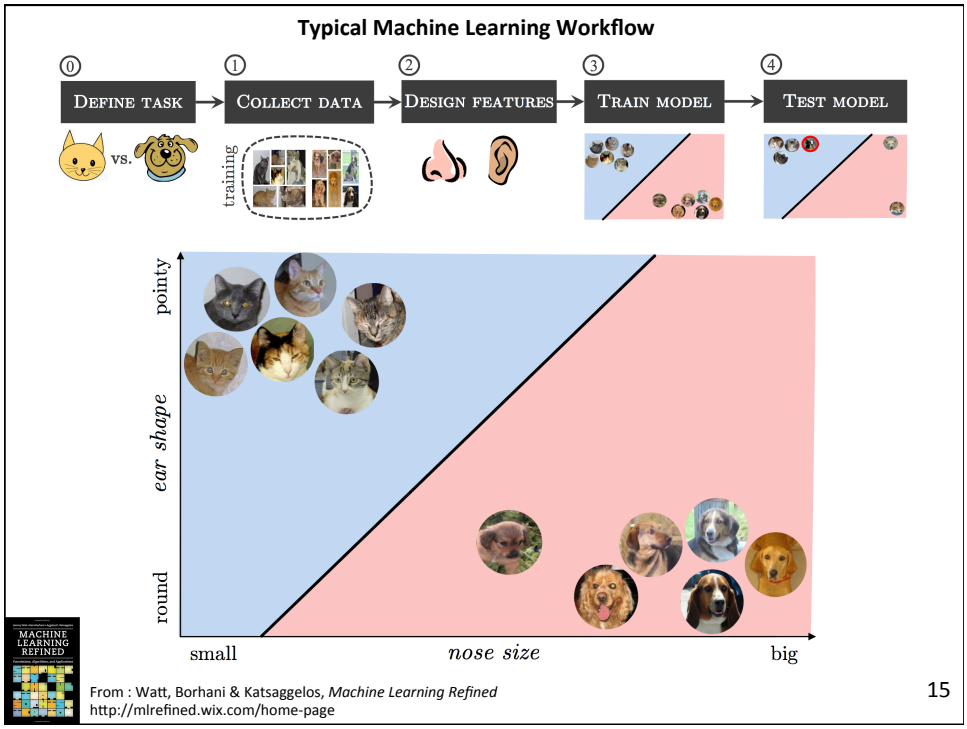
Typical Machine Learning Workflow





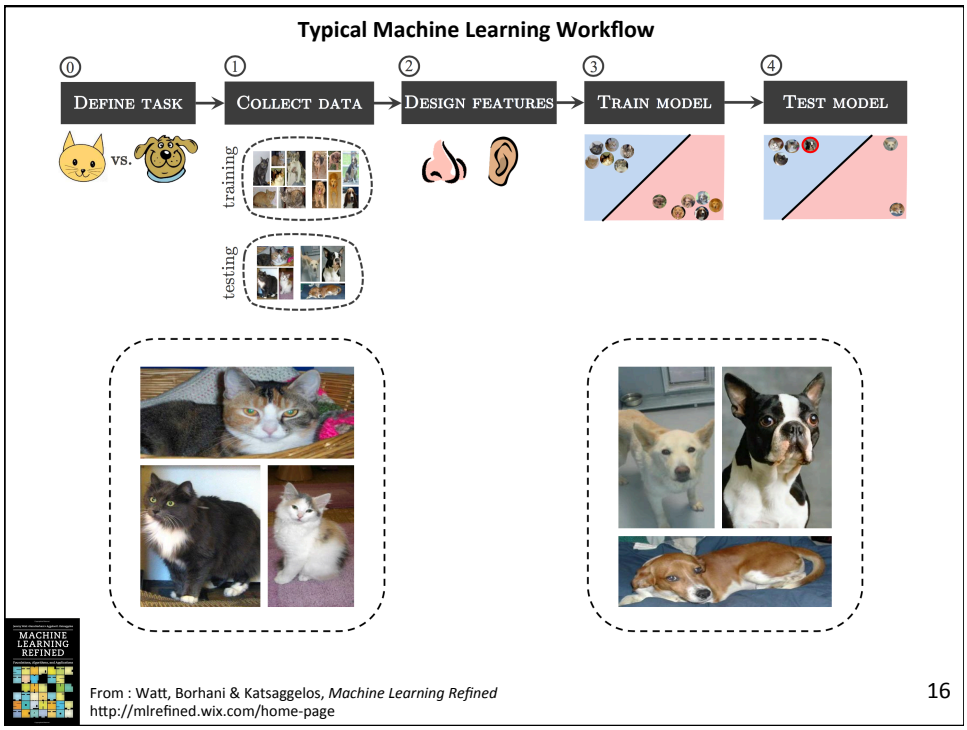
From : Watt, Borhani & Katsaggelos, *Machine Learning Refined*
http://mlrefined.wix.com/home-page

14



From : Watt, Borhani & Katsaggelos, *Machine Learning Refined*
http://mlrefined.wix.com/home-page

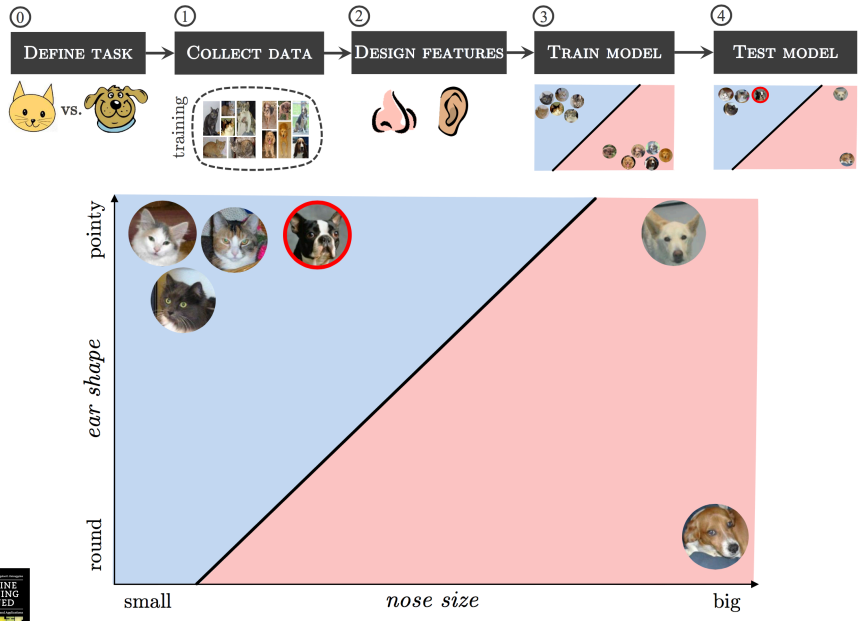
15



From : Watt, Borhani & Katsaggelos, *Machine Learning Refined*
http://mlrefined.wix.com/home-page

16

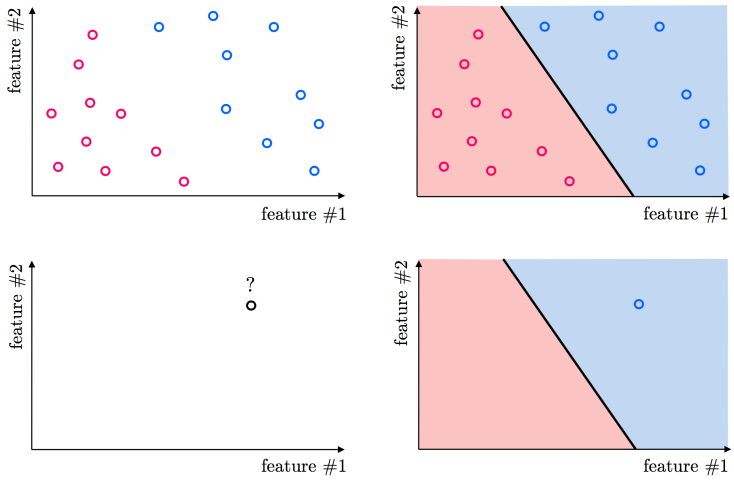
Typical Machine Learning Workflow



From : Watt, Borhani & Katsaggelos, *Machine Learning Refined*
http://mlrefined.wix.com/home-page

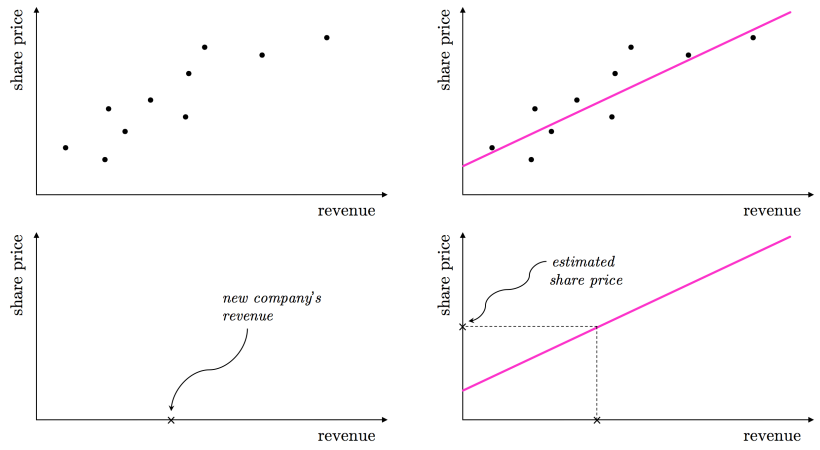
17

Classification



18

Regression



19

Three General Classes of ML

- **Supervised learning** – model $p(y|\mathbf{x})$
  - Given data and model, or data with correct output (label)
  - Regression, Classification, etc.
- **Unsupervised Learning** – model $p(\mathbf{x})$
  - Only given input data (no output)
  - Clustering, Latent Models, Projection methods, etc.
- **Reinforcement Learning** – model $p(\mathbf{s}_{t+1}|\mathbf{s}_t, a)$
  - Given input data, *some* output, and *grade* for output
  - Learning to choose better actions
  - Markov decision processes, POMDPs, planning

20

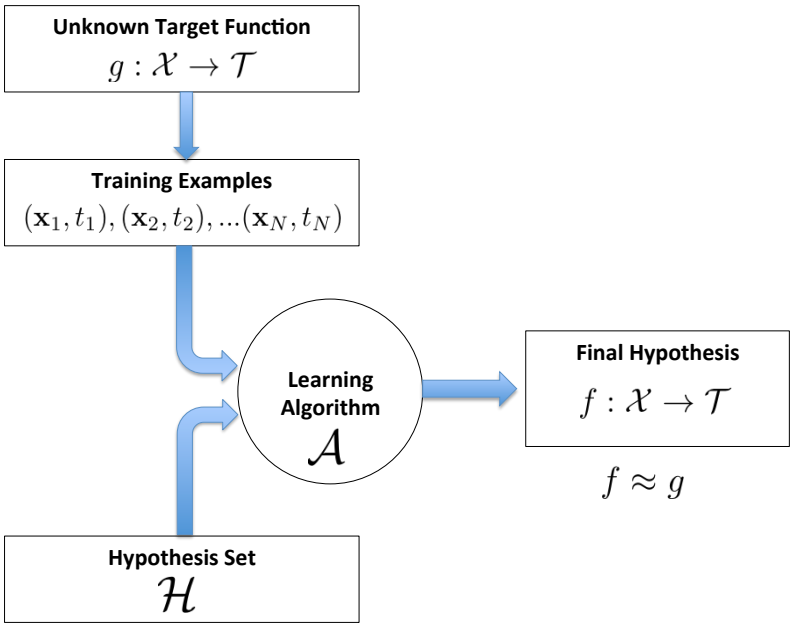
Some Terminology

- Input: $\mathbf{x}$ (customer application)
- Output: t (good/bad customer?)
- Target function: $g : \mathcal{X} \rightarrow \mathcal{T}$ (*ideal* credit approval fn)
- Data: $(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2), ...(\mathbf{x}_N, t_N)$ (historical records)

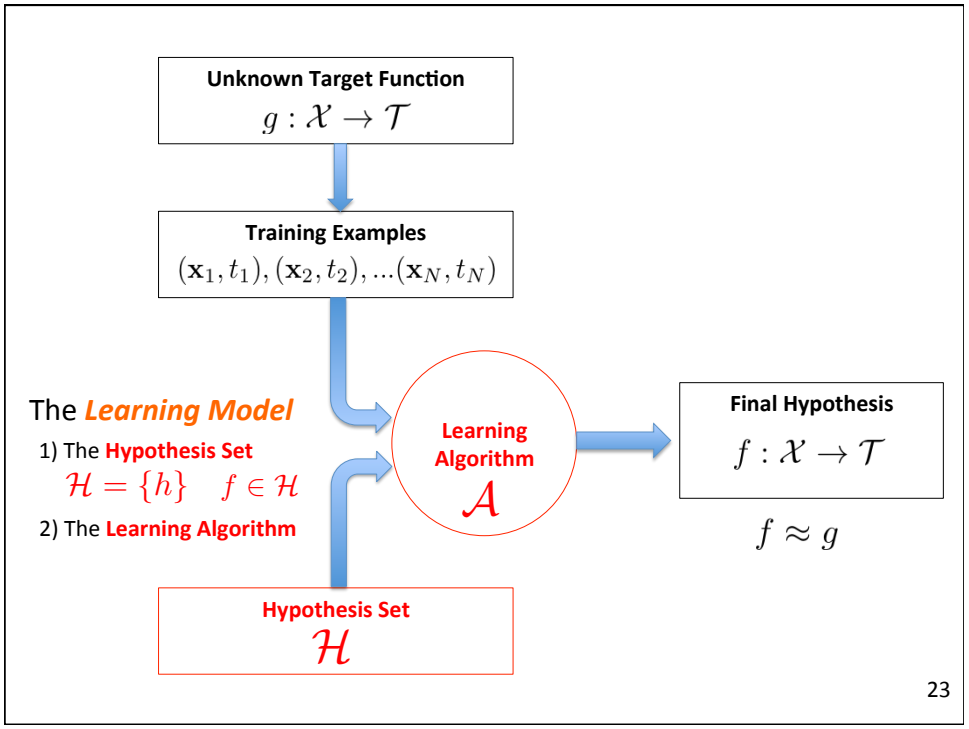
- Hypothesis: $f : \mathcal{X} \rightarrow \mathcal{T}$ (formula to be used)

Adapted from Yaser S. Abu-Mostafa et al., *Learning from Data*

21



22


Unknown Target Function
$g : \mathcal{X} \rightarrow \mathcal{T}$
Training Examples
$(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2), ...(\mathbf{x}_N, t_N)$
The *Learning Model*
1) The Hypothesis Set
$\mathcal{H} = \{h\} \quad f \in \mathcal{H}$
2) The Learning Algorithm
Learning Algorithm
$\mathcal{A}$
Final Hypothesis
$f : \mathcal{X} \rightarrow \mathcal{T}$
$f \approx g$
Hypothesis Set
$\mathcal{H}$
23