Do not distribute course material

Welcome to CS 4563

Thanks to:

- Some of the material from this course is from Prof. Sundeep Rangan
  - This includes many of the slides, the outline of the course, the demos, the programming assignments (aka labs) and many of the written homework assignments
- Some slides are from:
  - A. Zisserman, “Machine Learning Introduction”
  - Watt, Borhani, Katsaggelos, “Machine Learning Refined (Foundations, Algorithms, and Applications)

- Prof: Linda Sellie
    - Tuesdays & Thursdays 3:30 - 4:30, and by appointment. Office hours will be in person if I lecture in person. My office is room 848 on the 8th floor of 370 Jay Street
- The TAs for the course are MeiNa and Oliver. You can ask them questions on the Slack channel for the course or by attending office hours **Do not email the TAs directly.**
    - TA's office hours: To be determined
- Texts:
    - Standford notes https://cs229.stanford.edu/lectures-spring2022/main_notes.pdf
    - Deep Learning Chapter 5 https://www.deeplearningbook.org/contents/ml.html
    - *Hands-On Machine Learning with Scikit-Learn and TensorFlow* 2nd edition by Aurlien Gron. On line code https://github.com/ageron/handson-ml2
    - *Machine Learning for Humans* https://medium.com/machine-learning-for-humans/supervised-learning-740383a2feab

- Grading:
    - Midterm 1: 25%, midterm 2: 30%, end of semester quiz 10%; final project: 20%; Brightspace online quizzes: 5%; homework: 10%
      3% bonus points for reading/watching pre-lecture material, and participating in in-person lectures. Often the points will be based on your answer to Poll Everywhere questions.
    - Homework assignments typically will have a written part and a programming part. The programming parts (Labs) will involve python-based exercises (many of the assignments are modified versions of Prof. Sundeep Rangan's labs)

      Homework assignments will be turned in on Gradescope. The written portion must be typed into Gradescope or submitted. The TAs may send more information later

      Homework assignments that **are not** due just before an exam may be up to two days late, for a 10% deduction in your grade for each day. To make it easier to grade, typed homework assignments will be given a 5 point bonus
    - Final project is done in groups of two
    - Midterm exams and end of semester quiz are closed book. The online quizzes are open book, open notes. You must do your own work and you may have as many times a you would like before the deadline

- Pre-requisites:
    - One of: MA-UY 2224 (Data Analysis), MA-UY 2222 (Data Analysis 2) or EE-UY 2223 (Probability) or equivalent
    - Undergraduate probability and linear algebra (MA-UY 2034, MA-UY 2034G, MA-UY 3044 or MA-UY 3054)
    - Programming experience is essential. CS-UY 1134 (Data Structures and Algorithms)

We will be coding up our algorithms in Python!

The schedule is tentative and subject to change. We may not get through all the material in this schedule.

- Introduction
  - Course logistics. Examples of machine learning problems used today. Formulate machine learning problems (identify task, data, objectives). Classify ML problems as supervised vs. unsupervised, regression vs. classification.
- Linear regression
  - Least squares formula, Gradient Descent, Normal Equations Method
- Model selection and regularization: Identify the order in a multiple linear regression model
  - Understanding underfitting and overfitting with polynomials; irreducible error; bias and variance tradeoff; cross validation; regularization techniques
- Logistic Regression
  - Gradient ascent
- Support vector machines (SVMs)
  - Maximum margin; duality; kernel methods
- Neural networks
  - Formulation; back propagation
- Convolutional and deep networks
  - Convolutional layers; pooling layers
- PCA
  - Dimensionality reduction
- Clustering and K-Means
  - K-means; mixture models; EM methods
- TBA
- Final project presentations

Approach

- Understand the core algorithms of ML and how we can define them mathematically
- How to implement many of the algorithms from scratch
- Run many of the algorithms on real world datasets and evaluate how well they perform

Learning Objectives

- ❑Provide examples of machine learning used today
- ❑Classify a machine learning task:
  - Supervised vs. unsupervised, regression vs. classification
- ❑For supervised learning, identify the features/predictors and labels/target variables
- ❑Given a new problem, qualitatively describe how machine learning can be used
  - Formulate a potential machine learning task
  - Identify the data needed for the task
  - Identify objectives

Topic 1
What is Machine Learning?

INTRODUCTION TO MACHINE LEARNING

Outline

- ❑What is Machine Learning?
- ❑Types of machine learning algorithms
  - Supervised learning
    - Classification
    - Regression
  - Unsupervised learning
- ❑Why the hype today?

- “**Machine learning** is a field of computer science that gives computers the ability to learn without being *explicitly* programmed.”

  https://en.wikipedia.org/wiki/Machine_learning

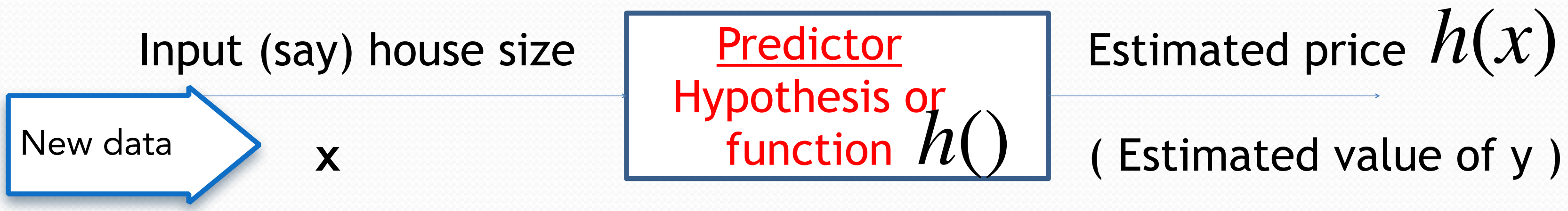
- “A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.” Tom M. Mitchell, CMU

Task	Performance Measure	Experience
Predict if email is spam	% of misclassified emails*	A set of spam and non-spam emails
Predict the change how much a stock price will change tomorrow	Squared difference between true price and predicted price	Historical stock prices
Read a handwritten digit	% of misclassified digits	A set of handwritten digits with the correct digit

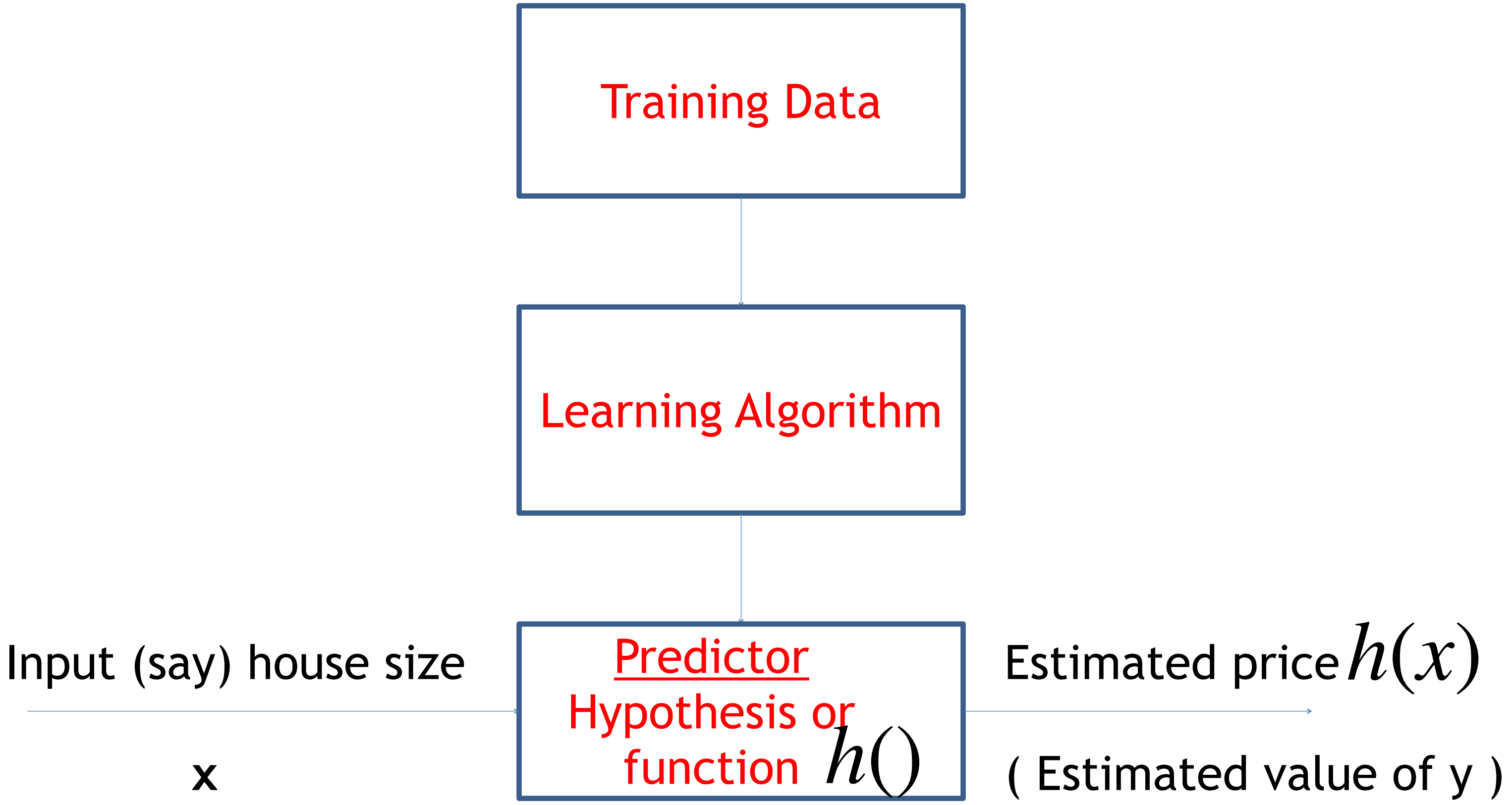
In-Class exercise
Identify:
Performance measure,
Experience

Why Machine Learning?

- Goal: Computers that “learn” from data or experience
- Where useful:
  - Human expertise does not exist (navigating on Mars)
  - Humans are unable to explain their expertise (speech recognition, driving)
  - Solution changes in time (routing on a computer network)
  - Solution should be customized to user (targeted advertising)



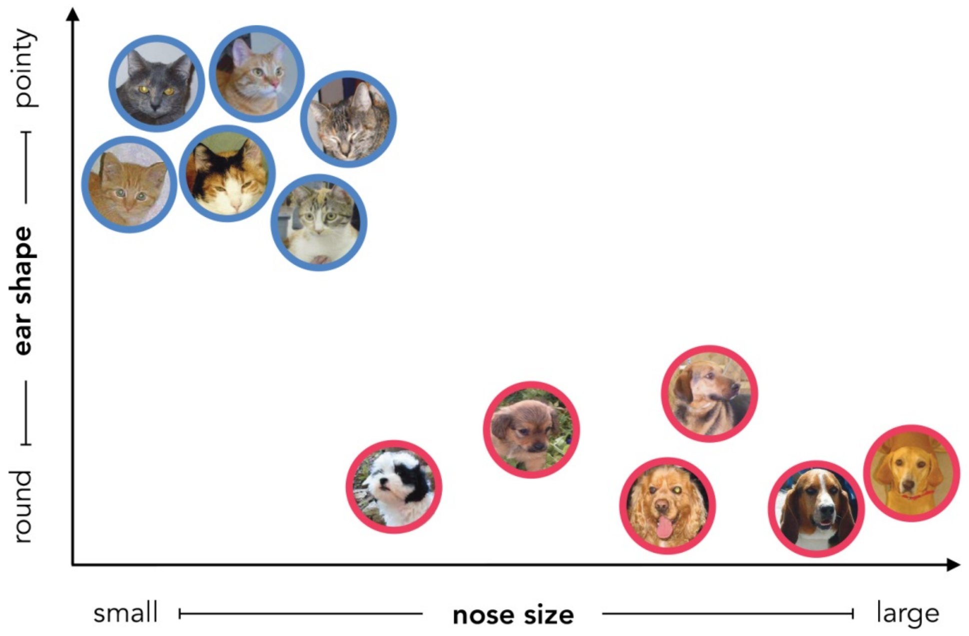
Adapted from by L. Hellerstein and L. Sellie for CS-GY 6923, NYU


Training Data
Learning Algorithm
Input (say) house size
x
Predictor
Hypothesis or
function $h()$
Estimated price $h(x)$
(Estimated value of y)

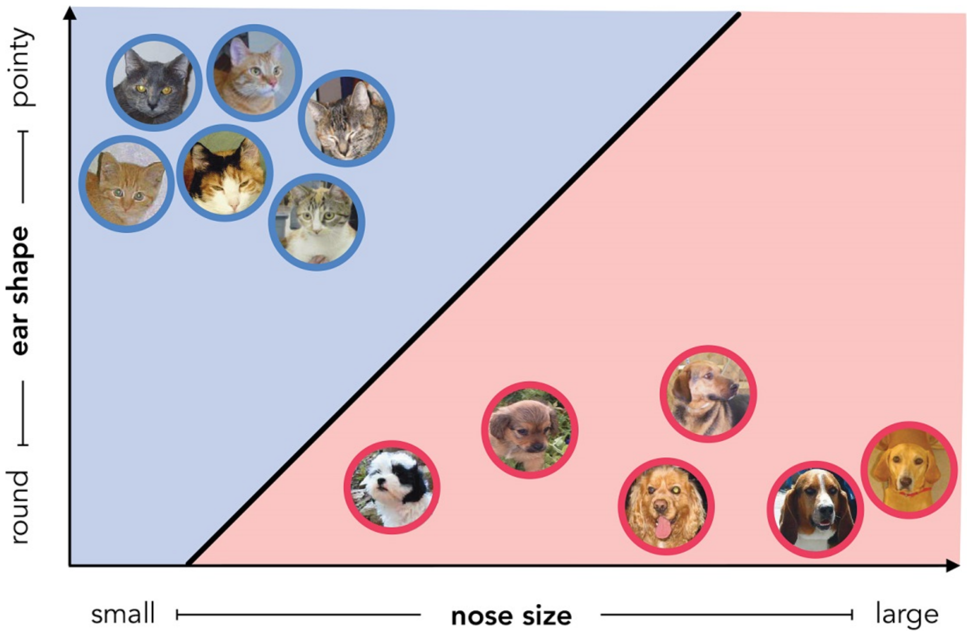
Watt Borhani and Katsaggelos_Fig1_01

Feature engineering

Watt Borhani and Katsaggelos_Fig1_02

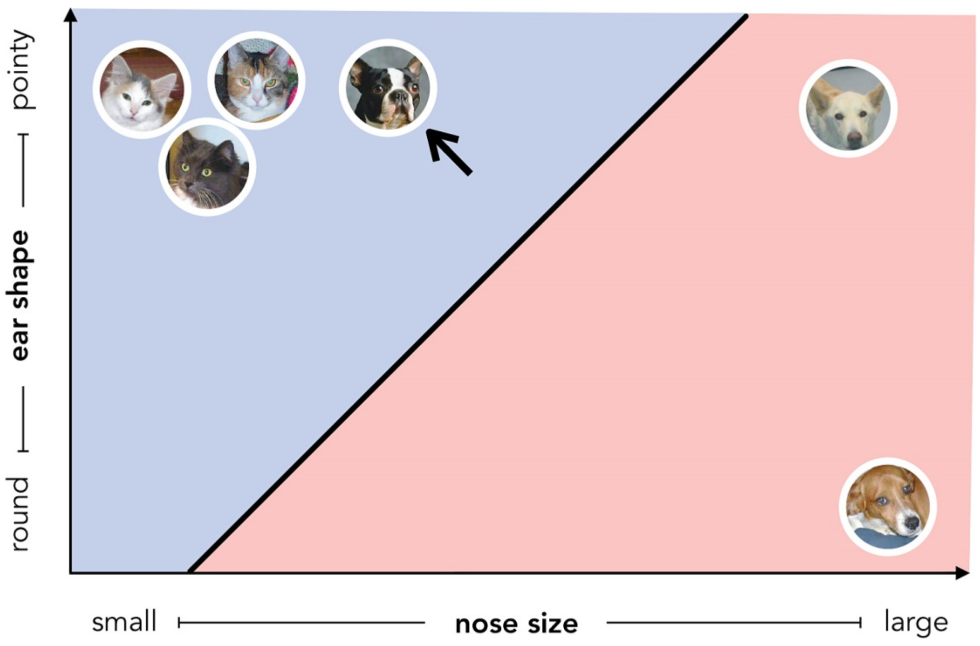


Watt Borhani and Katsaggelos_Fig1_03

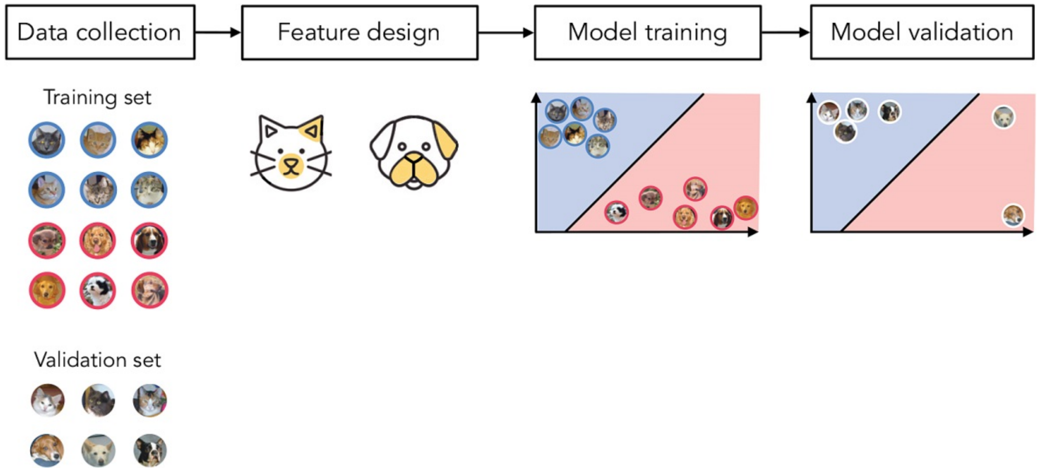


Watt Borhani and Katsaggelos_Fig1_04

Watt Borhani and Katsaggelos_Fig1_05



Watt Borhani and Katsaggelos_Fig1_06



Outline

- What is Machine Learning?
- Types of machine learning algorithms
  - Supervised learning
    - Classification
    - Regression
  - Unsupervised learning
- Why the hype today?

Types of learning we will cover

- Supervised Learning
  - Classification
  - Regression

  - Goal is to predict a specific value
  - Training examples have labels
  - Direct evaluation of accuracy
- Unsupervised Learning
  - Goal is to "understand" the data (structure/patterns)
  - Training examples don't have labels
  - Evaluation is less direct (qualitative or effect on another algorithm which can be measured directly)

Supervised learning goal:

Having observed a set of examples (input and output), create an rule for predicting the output for an unseen input

Supervised Learning (Continued)

Types of Supervised Learning:

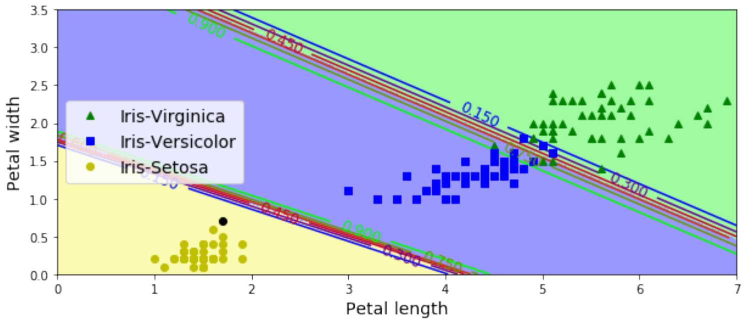
1. Regression: When the value to be predicted is a continuous/quantitative

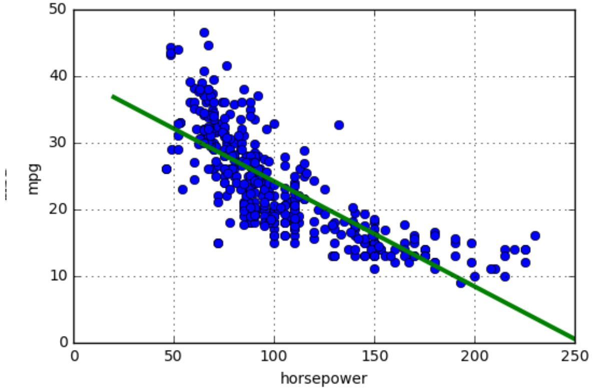
   e.g. predicting the map of a car, predicting price of house, predicting tomorrow's temperature in a weather predicting problem etc.

2. Classification: When the value to be predicted is a categorical/qualitative value

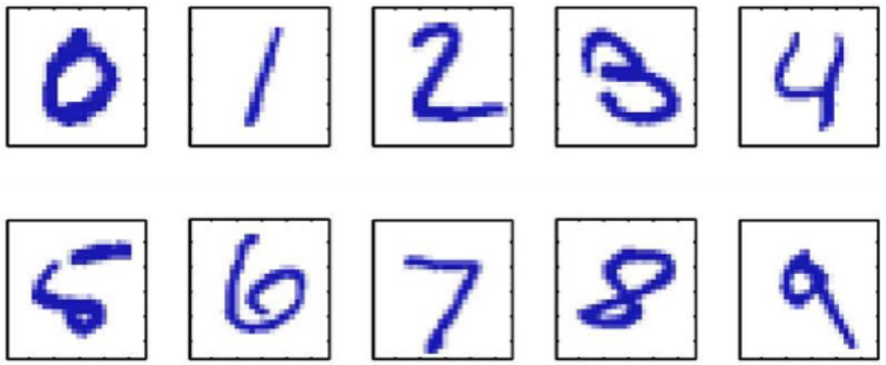
   e.g. which type of iris, whether a mail is spam(1) or not-spam (0), whether image contains a human face(1) or not (0) or whether a transaction is fraudulent(1) or not (0). This 1 or 0 is known as the class.

   There can be more than 2 categories. This is known as multi-class classification. E.g. whether an image contains apple(0) or orange (1) or mango (2).





Example: Digit Recognition



Images are 28 x 28 pixels

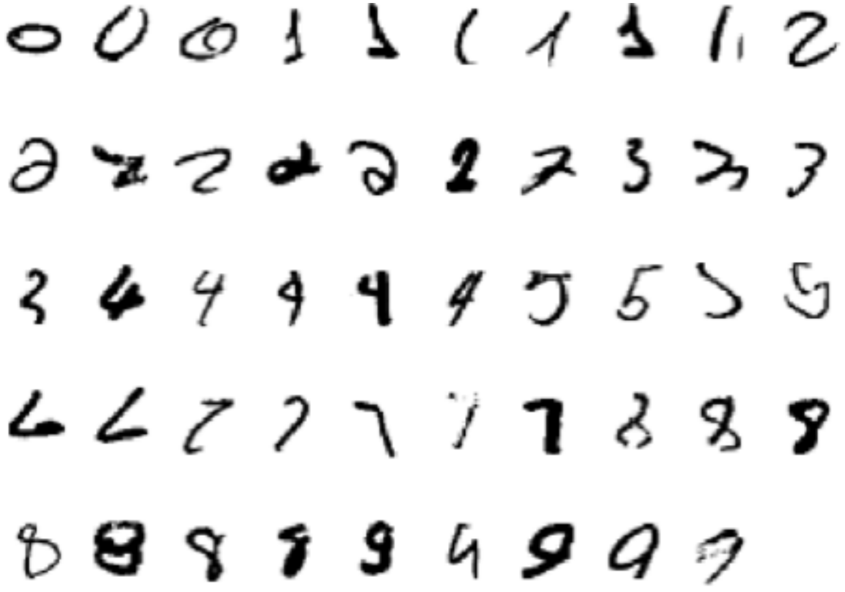
8x8 matrix for 0

$$\begin{bmatrix} 00 & 513 & 91 & 00 \\ 00 & 1315 & 1015 & 50 \\ 03 & 152 & 011 & 80 \\ 04 & 120 & 08 & 80 \\ 05 & 80 & 09 & 80 \\ 04 & 110 & 112 & 70 \\ 02 & 145 & 1012 & 00 \\ 00 & 613 & 100 & 00 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} 0 \\ 513 \\ 91 \\ 0 \\ 0 \\ \vdots \\ 613 \\ 100 \\ 0 \end{bmatrix}$$

- Recognize a digit from the image
- Learn a function $f(\boldsymbol{x}) \in \{0,1,\dots,9\}$, $\boldsymbol{x}$ is a 28 x 28 matrix
- Expert systems do not work well:
  - You can recognize the digits, but difficult to program a function $f(\boldsymbol{x})$ that works well
  - Try it!

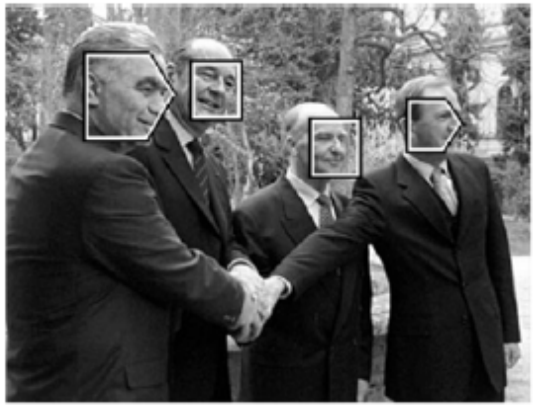
Supervised Learning

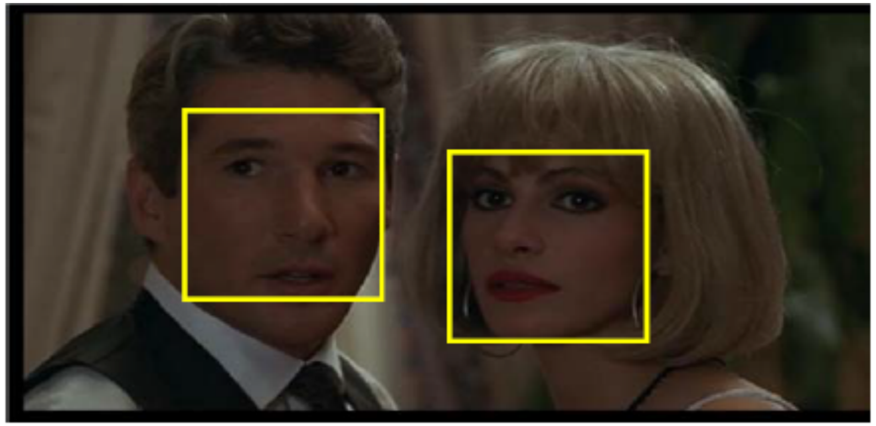
- Start with training data
- Ex: 6000 examples of each digit
- Learn a classifier that matches label well on training data
- Given new data use function to guess digit
- In 1/20/2018 systems got <0.21% errors http://rodrigob.github.io/are_we_there_yet/build/classification_datasets_results.html#4d4e495354
- First commercial application:
  - Used by USPS for recognizing zip codes on letters



Training examples
Each sample must be labeled by hand

Example: Face Detection

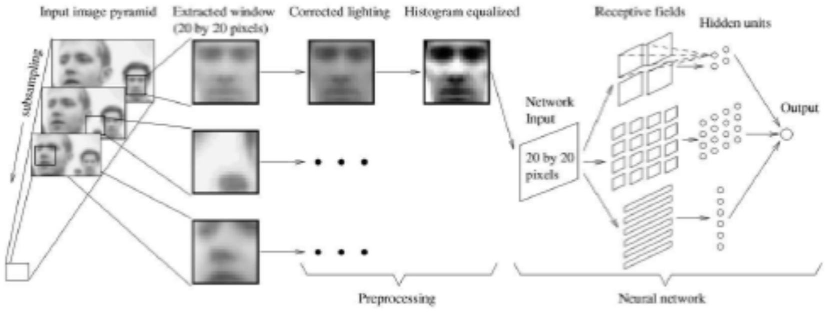




- Also a supervised learning problem
- For each image region, determine if
  - Face or non-face

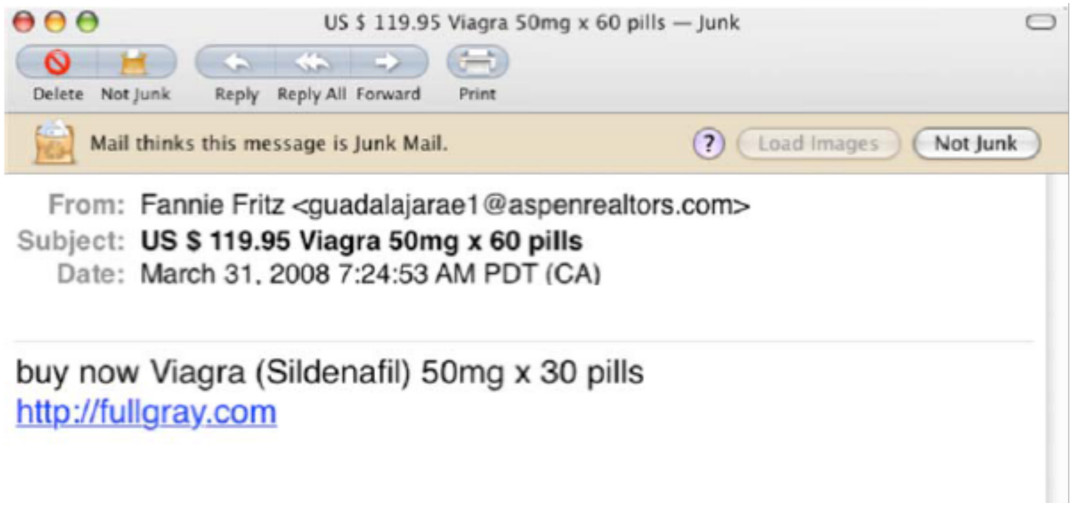
Face Recognition

- Typical early face recognition datasets:
- 5000 faces
  - All near frontal
  - Vary age, race, gender, lighting
- 10^8 non faces
- Faces are normalized (scale, translation)
- "functions" that work well may be very complex
- Many more datasets are available now:
  - See http://www.face-rec.org/databases/
  - You can use this for your project!



Rowley, Baluja and Kanade, 1998

Example: Spam Detection



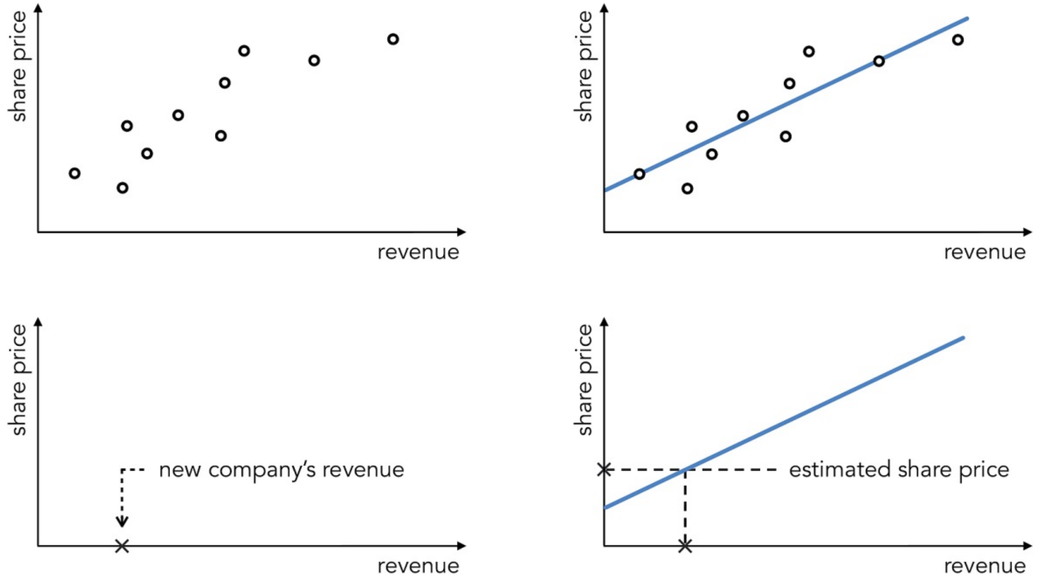
- Classification problem:
  - Is email junk or not junk?
- For ML, must represent email numerically
  - Common model: bag of words
  - Enumerate all words,
  - Represent email via word count num instances of word
- Challenge:
  - Very high-dimensional vector
  - System must continue to adapt (keep up with spammers)

Supervised Learning - Regression

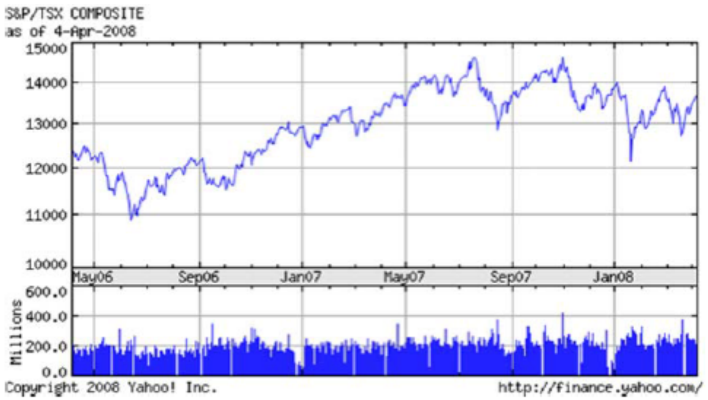
- Output (**label**/ target variable) is continuous (real-valued): y
- Input (**features**/ predictors): $x_1, x_2, \ldots, x_d$
  - e.g., predict tomorrow's temperature the features could be humidity level, wind speed, etc
- Function: machine learning algorithm learns a **hypothesis** (model/function) that it uses to predict
  - Linear regression learns a linear function $\hat{y} = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_d x_d$ by finding the values of $w_0, w_1, \cdots, w_d$

Notice the "hat".

Watt Borhani and Katsaggelos_Fig1_07



Example: Stock Price Prediction



- Can you predict the price of a stock?
- What variables would you use?

Supervised learning goal:

Input

Output

Given training examples

$$(\mathbf{x}^{(1)}, y^{(1)}), (\mathbf{x}^{(2)}, y^{(2)}), \ldots, (\mathbf{x}^{(N)}, y^{(N)})$$

Find: A good approximation to $f : \mathcal{X} \rightarrow \mathcal{Y}$

where $\mathcal{X}$ is the input space

$\mathcal{Y}$ is the output space

Examples:

digit recognition -

pixels -> {0,1,2,3, 4, 5,6,7,8,9}

face detection -

pixels -> {Face, Not a Face}

Spam detection-

email -> {Spam, Not Spam}

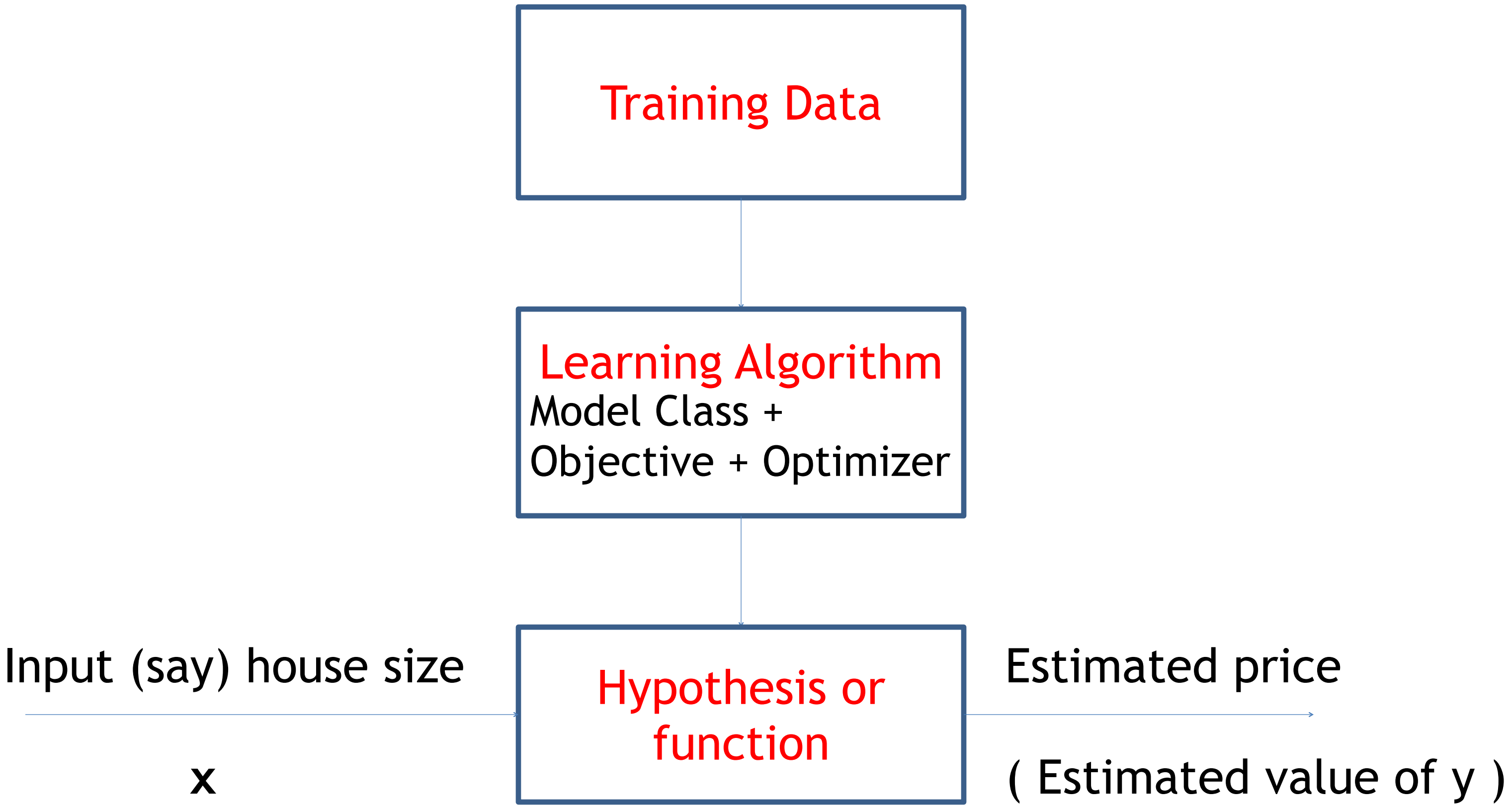
Classification problem: If the output is categorical/ quantitative

Stock Price Prediction-

historical prices, interest rate, etc -> $\mathbb{R}$

Regression problem: If the output is continuous/ quantitative

Supervised Learning



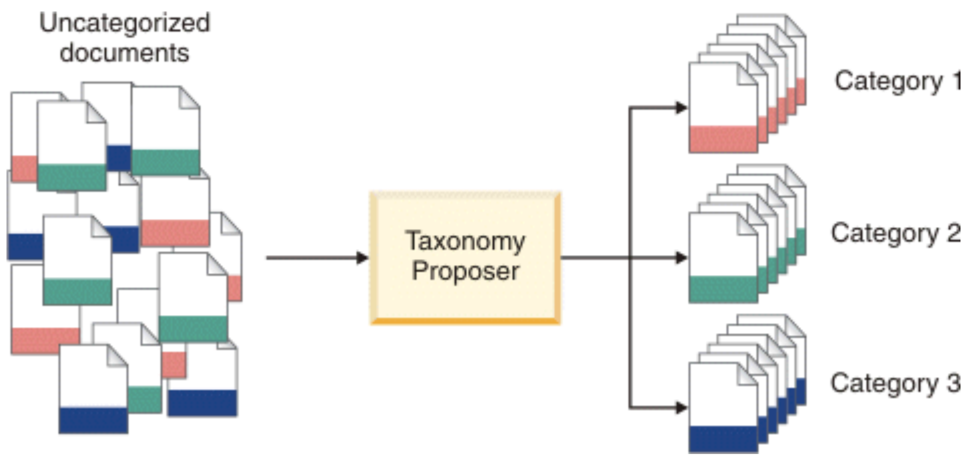
Unsupervised learning goal:

Having observed a set of examples (input only)

- group the examples into clusters
- Reduce the number of dimensions in the dataset

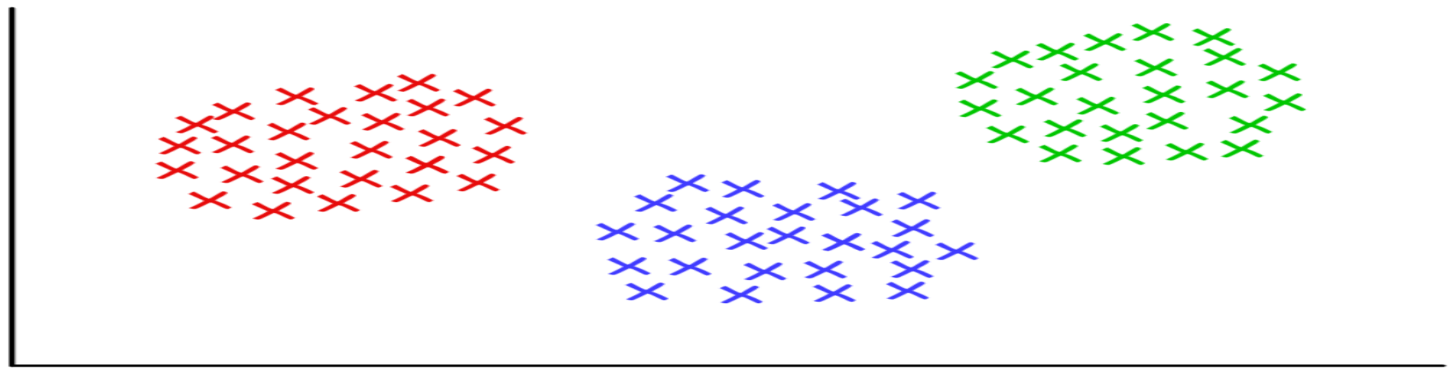
Unsupervised Learning

- Here no output variable or label is to be predicted.
- Only features are given and we need to find some pattern in data. E.g. say we are given some news articles and we need to cluster these articles into 2 clusters. This is known as clustering.
- We don’t know in advance whether which article belongs to which cluster.
- We just cluster them based on similarity in the features.
- Now, if we have a new article, we can tell by it’s text that which cluster it belongs to.
- There can be more than 2 clusters.



Example: Document classification
http://www.ibm.com/support/knowledgecenter
/SSBRAM_8.7.0/com.ibm.classify.ccenter.doc/
c_WBG_Taxonomy_Proposer.htm

Unsupervised Learning (continued)



- Some examples of clustering are: Organize computing clusters, social network analysis (e.g. grouping Facebook users based on their interest, or activities to target advertisements), market segmentation etc.

Watt Borhani and Katsaggelos_Fig1_14

There are many other types of machine that we will not cover

ML is a big field!

E.g.

- semi-supervised Learning
- reinforcement Learning

Outline

- ❑What is Machine Learning?
- ❑Types of machine learning algorithms
  - Supervised learning
    - Classification
    - Regression
  - Unsupervised learning
- ❑Why the hype today?

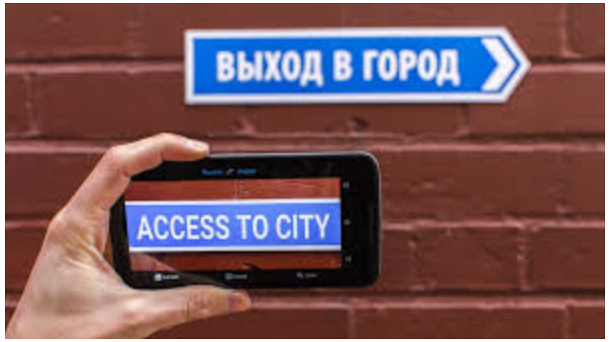
Machine Learning in Many Fields

- Retail: Market basket analysis, Customer relationship management (CRM)
- Finance: Credit scoring, fraud detection
- Manufacturing: Control, robotics, troubleshooting
- Medicine: Medical diagnosis
- Telecommunications: Spam filters, intrusion detection
- Bioinformatics: Motifs, alignment
- Web mining: Search engines
- ...

What ML is Doing Today?

- Autonomous driving
- Jeopardy
- Very difficult games: Alpha Go
- Machine translation

- Many, many others…



Why Machine Learning is in Hype from the past few years?

- There are 3 main factors behind recent increased applications of ML:

1. **Data**: There has been a huge amount of data generated in the past decade and continuously increasing due to internet, massive storage, digitalization etc.
2. **Computation Power**: Hardware advance like GPU's, distributed machines etc.
3. **Algorithmic Innovation**: New and optimized algorithms by ML and AI research Community.

Journals

- Journal of Machine Learning Research www.jmlr.org
- Machine Learning
- Neural Computation
- Neural Networks
- IEEE Trans on Neural Networks and Learning Systems
- IEEE Trans on Pattern Analysis and Machine Intelligence
- Journals on Statistics/Data Mining/Signal Processing/Natural Language Processing/ Bioinformatics/...

Conferences

- International Conference on Machine Learning (ICML)
- European Conference on Machine Learning (ECML)
- Neural Information Processing Systems (NIPS)
- Uncertainty in Artificial Intelligence (UAI)
- Computational Learning Theory (COLT)
- International Conference on Artificial Neural Networks (ICANN)
- International Conference on AI & Statistics (AISTATS)
- International Conference on Pattern Recognition (ICPR)

Exercise

- Break into small groups
- Take a field that interests you:
  - Ex. Driving a car, understanding social networks, recommend a movie to watch, …
- Identify a specific task that can be done with machine learning
  - What is the objective of the task?
  - What is the data you need?
  - What type of ML problem is this? Classification, regression, …
  - How would your approach compare to an expert-driven method?