

Introduction to Machine Learning

What is Learning?

- *“Learning denotes changes in a system that ... enable a system to do the same task ... more efficiently the next time.”* - Herbert Simon
- *“Learning is constructing or modifying representations of what is being experienced.”* - Ryszard Michalski
- *“Learning is making useful changes in our minds.”* - Marvin Minsky

“Machine learning refers to a system capable of the autonomous acquisition and integration of knowledge.”

Why “Learn”?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience.
- There is no need to “learn” to calculate payroll
- Learning is used when:
 - Human expertise does not exist (navigating on Mars),
 - Humans are unable to explain their expertise (speech recognition)
 - Solution changes in time (routing on a computer network)
 - Solution needs to be adapted to particular cases (user biometrics)

What We Talk About When We Talk About “Learning”

- Learning general models from a data of particular examples
- Data is cheap and abundant (data warehouses, data marts); knowledge is expensive and scarce.
- Example in retail: Customer transactions to consumer behavior:
People who bought “Da Vinci Code” also bought “The Five People You Meet in Heaven” (www.amazon.com)
- Build a model that is **a good and useful approximation** to the data.

So What Is Machine Learning?

- Automating automation
- Getting computers to program themselves
- Writing software is the bottleneck
- Let the data do the work instead!

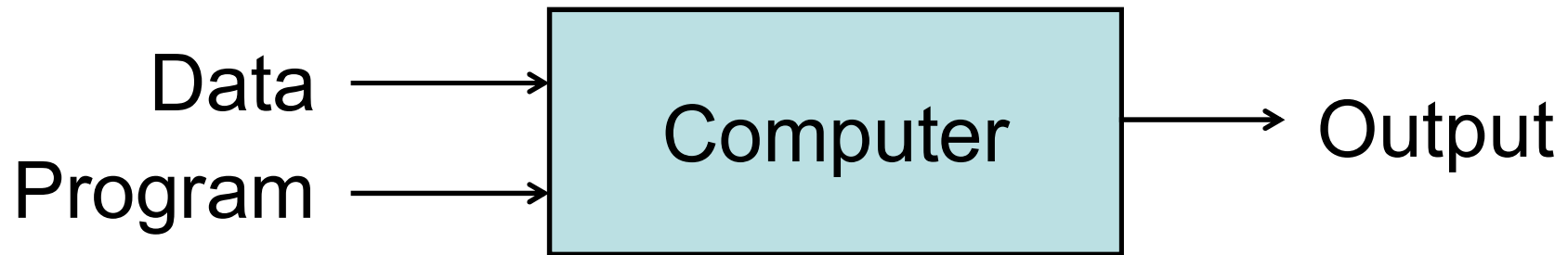
What is Machine Learning?

- Machine Learning
- Study of algorithms that improve their performance at some task with experience
- Optimize a performance criterion using example data or past experience.
- Role of Statistics: Inference from a sample
- Role of Computer science: Efficient algorithms to
 - Solve the optimization problem
 - Representing and evaluating the model for inference

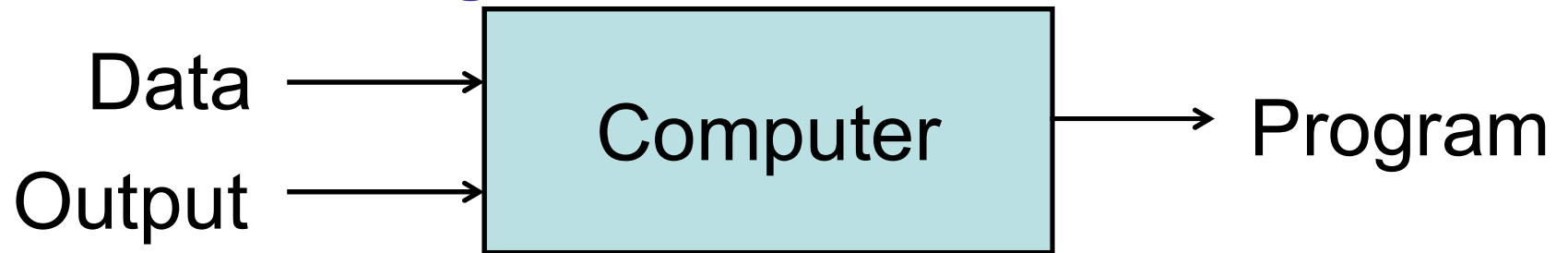
Why Machine Learning?

- No human experts
- industrial/manufacturing control
- mass spectrometer analysis, drug design, astronomic discovery
- Black-box human expertise
- face/handwriting/speech recognition
- driving a car, flying a plane
- Rapidly changing phenomena
- credit scoring, financial modeling
- diagnosis, fraud detection
- Need for customization/personalization
- personalized news reader
- movie/book recommendation

Traditional Programming



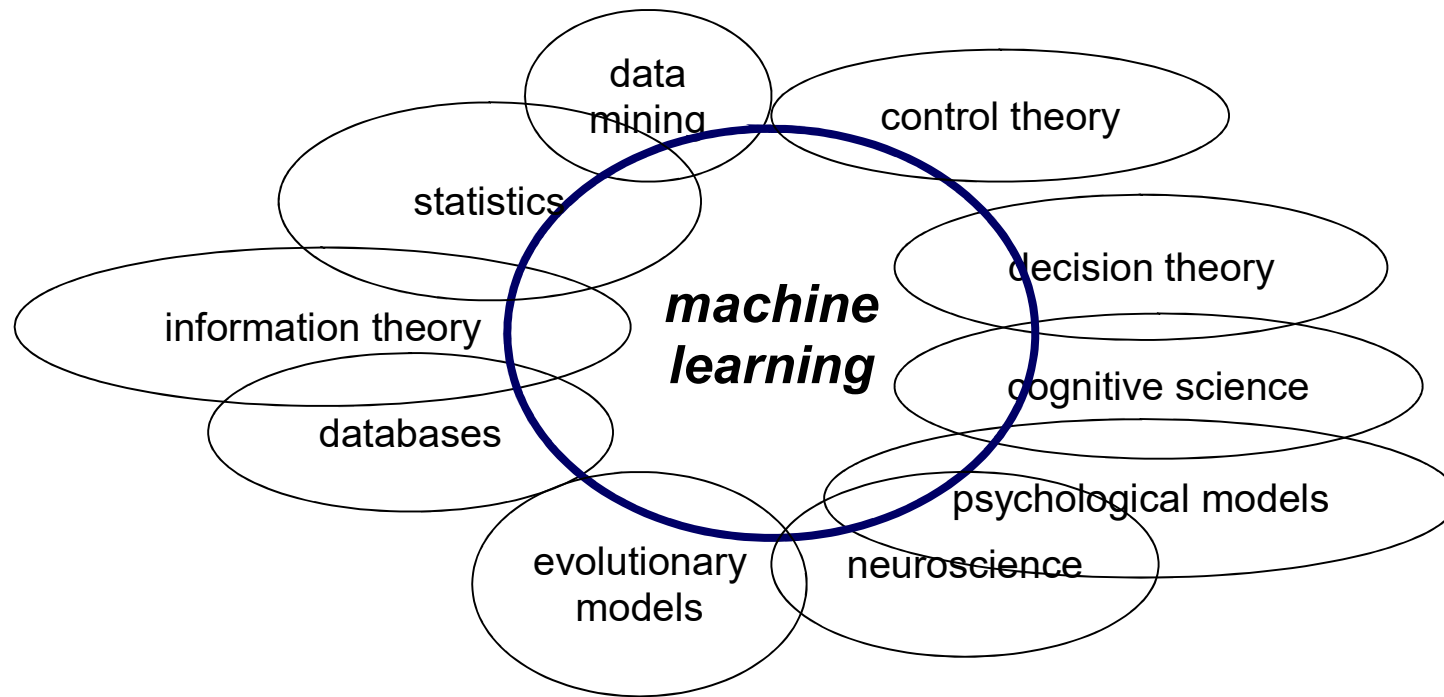
Machine Learning



Growth of Machine Learning

- Machine learning is preferred approach to
- Speech recognition, Natural language processing
- Computer vision
- Medical outcomes analysis
- Robot control
- Computational biology
- This trend is accelerating
- Improved machine learning algorithms
- Improved data capture, networking, faster computers
- Software too complex to write by hand
- New sensors / IO devices
- Demand for self-customization to user, environment
- It turns out to be difficult to extract knowledge from human experts → *failure of expert systems in the 1980's.*

Related Fields



Machine learning is primarily concerned with the accuracy and effectiveness of the *computer system*.

Concept Learning

- Inducing general functions from specific training examples is a main issue of machine learning.
- Concept Learning: Acquiring the definition of a general category from given sample positive and negative training examples of the category.
- Concept Learning can be seen as a problem of searching through a predefined space of potential hypotheses for the hypothesis that best fits the training examples.
- The hypothesis space has a general-to-specific ordering of hypotheses, and the search can be efficiently organized by taking advantage of a naturally occurring structure over the hypothesis space.

A Formal Definition for Concept Learning: Inferring a boolean-valued function from training examples of its input and output.

- An example for concept-learning is the learning of bird-concept from the given examples of birds (positive examples) and non-birds (negative examples).
- We are trying to learn the definition of a concept from given examples.

The Inductive Learning Hypothesis

- Although the learning task is to determine a hypothesis h identical to the target concept cover the entire set of instances X , the only information available about c is its value over the training examples.
 - Inductive learning algorithms can at best guarantee that the output hypothesis fits the target concept over the training data.
 - Lacking any further information, our assumption is that the best hypothesis regarding unseen instances is the hypothesis that best fits the observed training data. This is the fundamental assumption of inductive learning.
- The Inductive Learning Hypothesis - Any hypothesis found to approximate the target function well over a sufficiently large set of training examples will also approximate the target function well over other unobserved examples.

Dimensions of Learning Systems

- type of feedback
 - supervised (labeled examples)
 - unsupervised (unlabeled examples)
 - reinforcement (reward)
- representation
 - attribute-based (feature vector)
 - relational (first-order logic)
- use of knowledge
 - empirical (knowledge-free)
 - analytical (knowledge-guided)

ML in a Nutshell

- Tens of thousands of machine learning algorithms
- Hundreds new every year
- Every machine learning algorithm has three components:
 - **Representation**
 - **Evaluation**
 - **Optimization**

Representation

- Decision trees
- Sets of rules / Logic programs
- Instances
- Graphical models (Bayes/Markov nets)
- Neural networks
- Support vector machines
- Model ensembles
- Etc.

Evaluation

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence
- Etc.

Optimization

- Combinatorial optimization
 - E.g.: Greedy search
- Convex optimization
 - E.g.: Gradient descent
- Constrained optimization
 - E.g.: Linear programming

Types of Learning

- **Supervised (inductive) learning**
 - Training data includes desired outputs
- **Unsupervised learning**
 - Training data does not include desired outputs
- **Semi-supervised learning**
 - Training data includes a few desired outputs
- **Reinforcement learning**
 - Rewards from sequence of actions

Machine Learning Types

Supervised Learning

Unsupervised Learning

Semi-supervised learning

Reinforcement learning

Regression

Classification

Clustering

Clustering / Classification

Classification / Control

(e.g.)
Taxi Fare/
Housing price
prediction

(e.g.)
Sentiment/
Spam filter
prediction

(e.g.)
Product
Customer
Segmentation

(e.g.)
speech
analysis

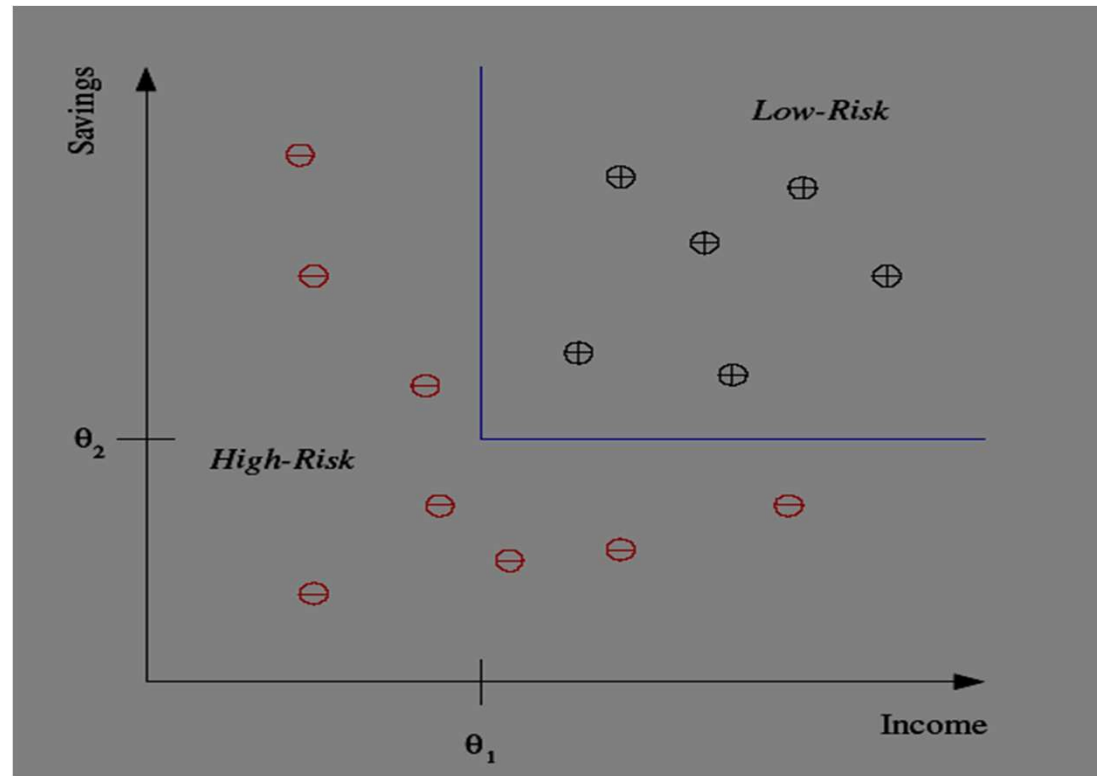
(e.g.)
Self-Driving
Car

Inductive Learning

- **Given** examples of a function $(X, F(X))$
- **Predict** function $F(X)$ for new examples X
- Discrete $F(X)$: Classification
- Continuous $F(X)$: Regression
- $F(X) = \text{Probability}(X)$: Probability estimation

Classification

- Example: Credit scoring
- Differentiating between **low-risk** and **high-risk** customers from their *income* and *savings*



Discriminant: IF $income > \theta_1$ AND $savings > \theta_2$
THEN **low-risk** ELSE **high-risk**

Model

Classification: Applications

- Aka Pattern recognition
- Face recognition:** Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition:** Different handwriting styles.
- Speech recognition:** Temporal dependency.
- Use of a dictionary or the syntax of the language.
- Sensor fusion: Combine multiple modalities; eg, visual (lip image) and acoustic for speech
- Medical diagnosis:** From symptoms to illnesses
- Web Advertizing:** Predict if a user clicks on an ad on the Internet.

Face Recognition

Training examples of a person



Test images



AT&T Laboratories, Cambridge UK
<http://www.uk.research.att.com/facedatabase.html>

Prediction: Regression

- Example: Price of a used car

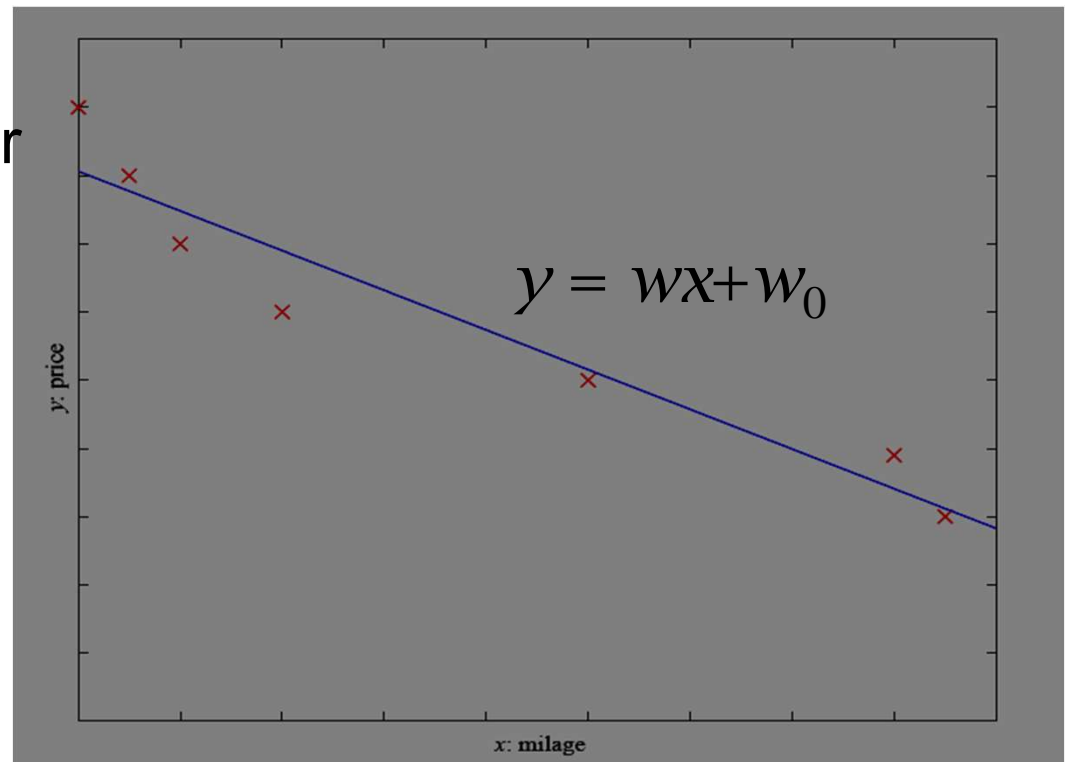
- x : car attributes

y : price

$$y = g(x | \theta)$$

$g(\cdot)$ model,

□ θ parameters



Supervised Learning: Uses

Example: decision trees tools that create rules

- **Prediction of future cases:** Use the rule to predict the output for future inputs
- **Knowledge extraction:** The rule is easy to understand
- **Compression:** The rule is simpler than the data it explains
- **Outlier detection:** Exceptions that are not covered by the rule, e.g., fraud

Unsupervised Learning

- Learning “what normally happens”
- No output
- Clustering: Grouping similar instances
- Other applications: Summarization, Association Analysis
- Example applications
 - Customer segmentation in CRM
 - Image compression: Color quantization
 - Bioinformatics: Learning motifs

Reinforcement Learning

- Topics:
- Policies: what actions should an agent take in a particular situation
- Utility estimation: how good is a state (→used by policy)
- No supervised output but delayed reward
- Credit assignment problem (what was responsible for the outcome)
- Applications:
- Game playing
- Robot in a maze
- Multiple agents, partial observability, ...

ML in Practice

1. Understanding domain, prior knowledge, and goals
2. Data integration, selection, cleaning, pre-processing, etc.
3. Learning models
4. Interpreting results
5. Consolidating and deploying discovered knowledge
6. Loop

Summary

- Introductory course that covers a wide range of machine learning techniques—from basic to state-of-the-art.
- You will learn about the methods you heard about: Naïve Bayes', regression, nearest-neighbor (kNN), decision trees, support vector machines, learning ensembles, over-fitting, regularization, dimensionality reduction & PCA, parameter estimation, mixture models, comparing models, density estimation, clustering centering on K-means, EM, and DBSCAN, active and reinforcement learning.
- Covers algorithms, theory and applications

What We'll Cover

- **Supervised learning**

- Decision tree induction
- Naïve Bayes'
- Regression
- nearest-neighbor (kNN)
- Over-fitting, Regularization,
- parameter estimation
- comparing models
- Neural networks
- Model ensembles

- **Unsupervised learning**

Clustering - clustering centering on K-means, EM, and DBSCAN,
Dimensionality reduction & PCA,
active and reinforcement learning.

It's going to be fun and hard work