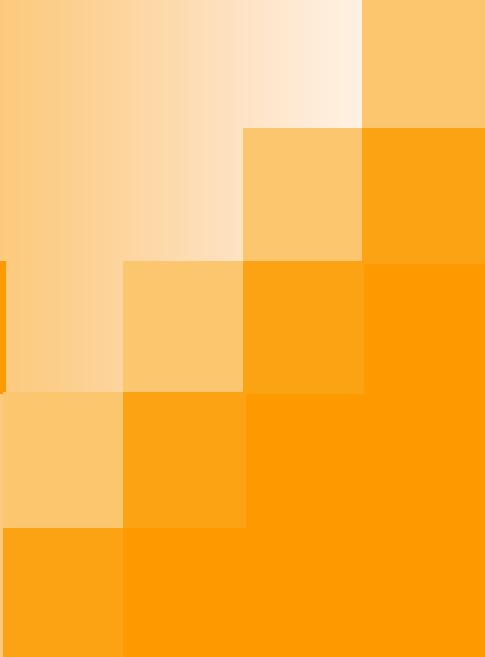


Machine Learning

Introduction

*Dr. Dipak Kumar Mohanty
KIIT Deemed to be University,
Bhubaneswr.*



CHAPTER 1:

Introduction

Source Materials

- G. Gopal. **Applied Machine Learning**
- R. Duda, P. Hart & D. Stork, ***Pattern Classification*** (2nd ed.), Wiley (Required)
- T. Mitchell, ***Machine Learning***, McGraw-Hill (Recommended)
- Papers

A Few Quotes

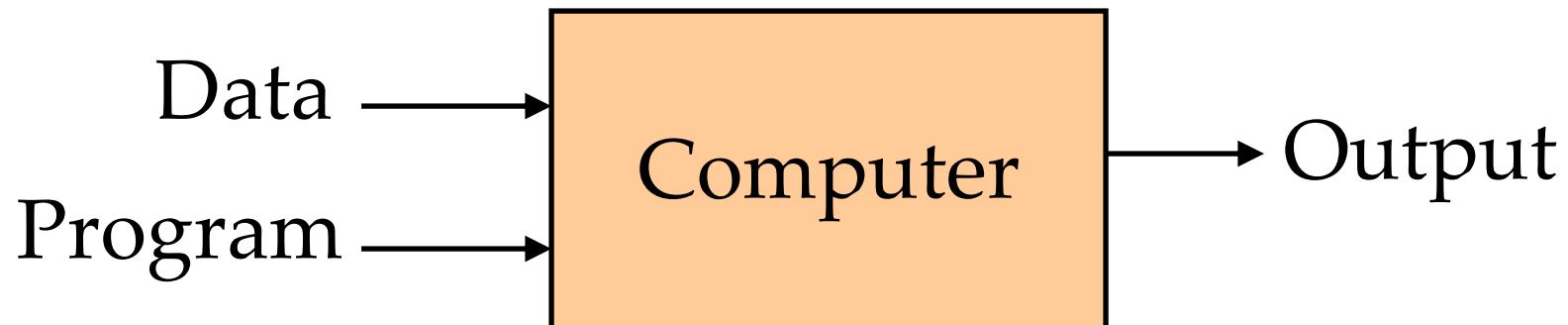
- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)
- “Machine learning is the next Internet”
(Tony Tether, Director, DARPA)
- “Machine learning is the hot new thing”
(John Hennessy, President, Stanford)
- “Web rankings today are mostly a matter of machine learning” (Prabhakar Raghavan, Dir. Research, Yahoo)
- “Machine learning is going to result in a real revolution” (Greg Papadopoulos, CTO, Sun)
- “Machine learning is today’s discontinuity”
(Jerry Yang, CEO, Yahoo)



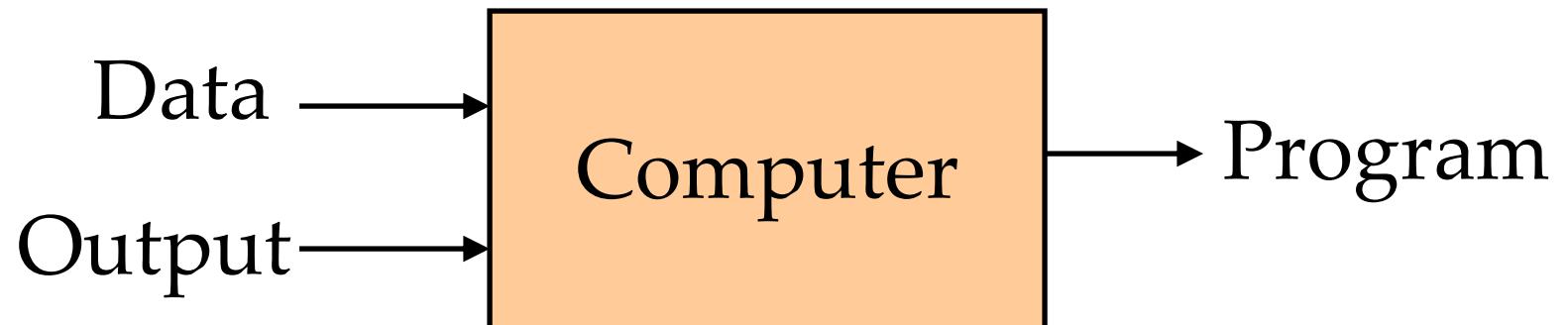
So What Is Machine Learning?

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

Traditional Programming



Machine Learning



Magic?

No, more like
gardening

- n **Seeds** = Algorithms
- n **Nutrients** = Data
- n **Gardener** = You
- n **Plants** = Programs





Sample Applications

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- Robotics
- Information extraction
- Social networks
- Debugging
- [Your favorite area]

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 (labelled input and classification, Regression)
 - Ex: Flowers, Exit poll
- **Unsupervised learning**
 - Training data does not include desired outputs(unlabelled data)
 - Ex: PK movie, K-mean Clustering
- **Semi-supervised learning**
 - Training data includes a few desired outputs
- **Reinforcement learning**
 - Rewards from sequence of actions
 - Ex: Dog, Baby learns

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

ML in Practice

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

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.

Data Mining/KDD

Definition := “*KDD is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data*” (Fayyad)

Applications:

- **Retail:** Market basket analysis, Customer relationship management (CRM)
- **Finance:** Credit scoring, fraud detection
- **Manufacturing:** Optimization, troubleshooting
- **Medicine:** Medical diagnosis
- **Telecommunications:** Quality of service optimization
- **Bioinformatics:** Motifs, alignment
- **Web mining:** Search engines
- ...

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

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.*

Applications

- Association Analysis
- Supervised Learning
 - Classification
 - Regression/Prediction
- Unsupervised Learning
- Reinforcement Learning

Learning Associations

- Basket analysis:

$P(Y | X)$ probability that somebody who buys X also buys Y where X and Y are products/services.

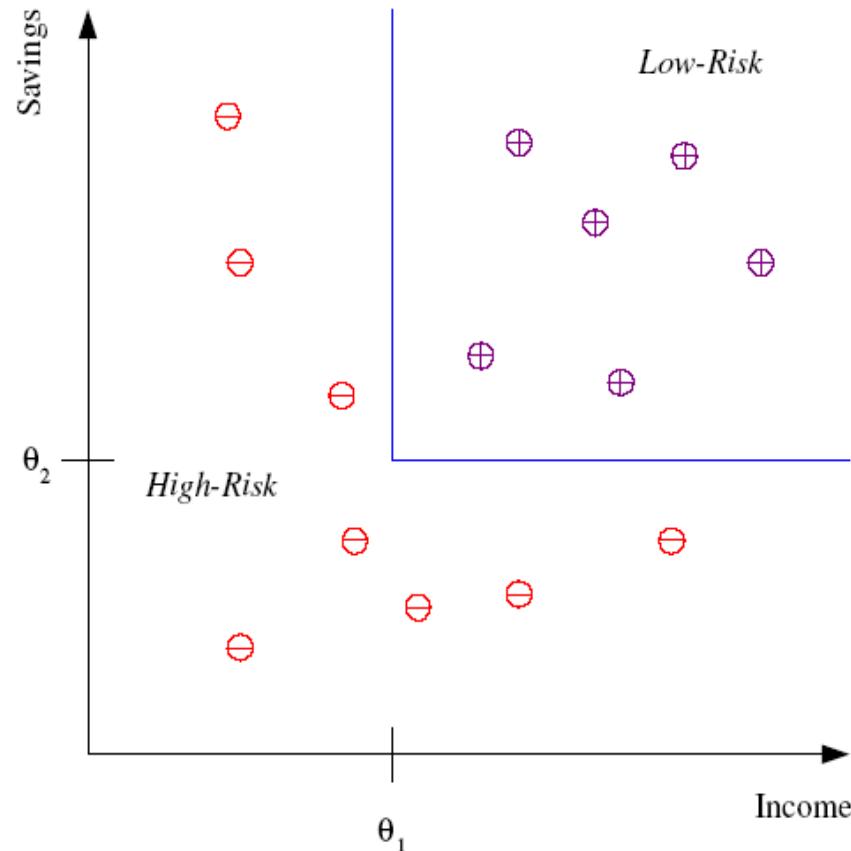
Example: $P(\text{ chips} | \text{ beer}) = 0.7$

Market-Basket transactions

<i>TID</i>	<i>Items</i>
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

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 Advertising: Predict if a user clicks on an ad on the Internet.

Face Recognition

Training examples of a person



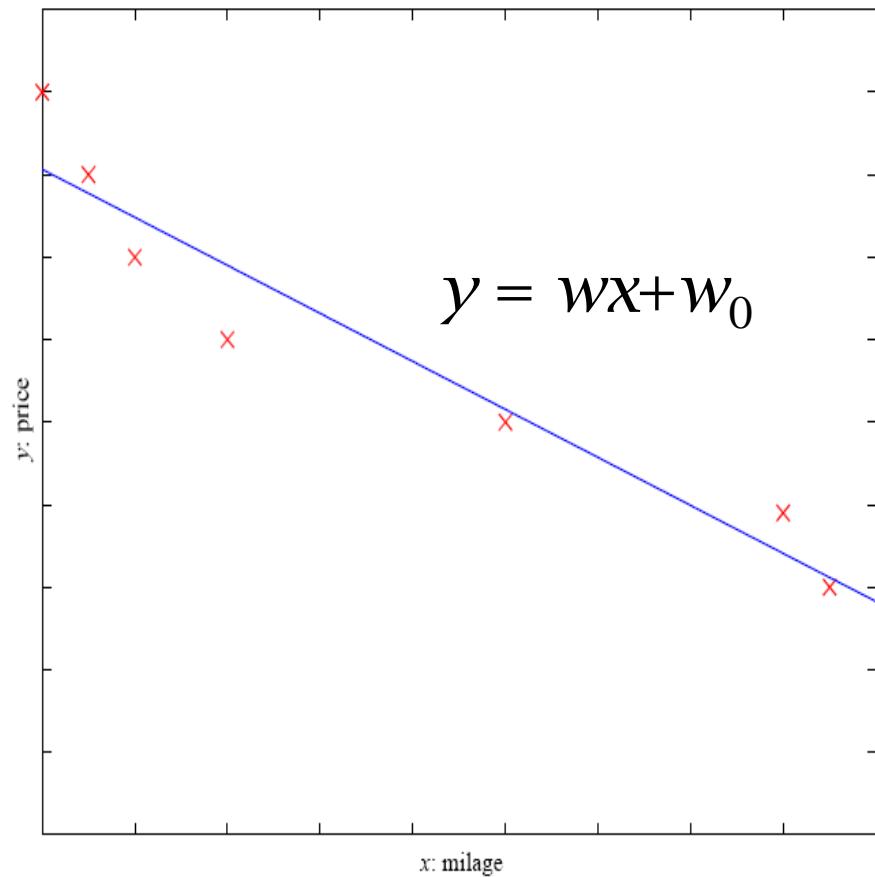
Test images



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

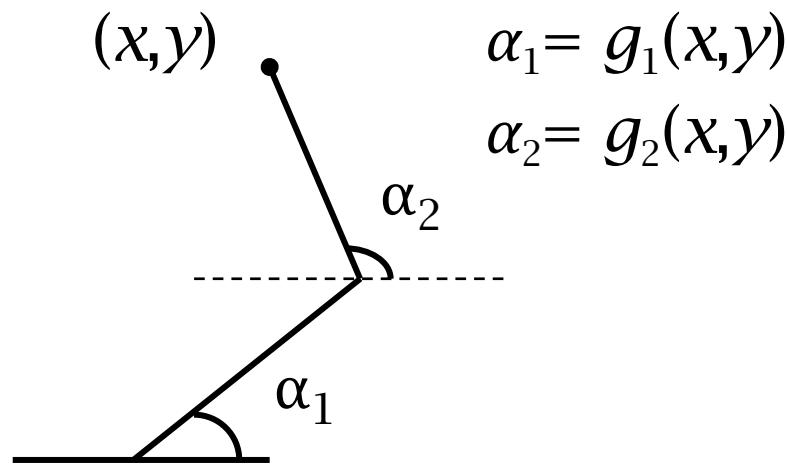
Prediction: Regression

- Example: Price of a used car
 - x : car attributes
 y : price
- $$y = g(x | \theta)$$
- $g(\cdot)$ model,
 θ parameters



Regression Applications

- Navigating a car: Angle of the steering wheel (CMU NavLab)
- Kinematics of a robot arm



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 (\rightarrow 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, ...

Resources: Datasets

- UCI Repository:
<http://www.ics.uci.edu/~mlearn/MLRepository.html>
- UCI KDD Archive:
<http://kdd.ics.uci.edu/summary.data.application.html>
- Statlib: <http://lib.stat.cmu.edu/>
- Delve: <http://www.cs.utoronto.ca/~delve/>

Resources: Journals

- Journal of Machine Learning Research www.jmlr.org
- Machine Learning
- IEEE Transactions on Neural Networks
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- Annals of Statistics
- Journal of the American Statistical Association
- ...

Resources: Conferences

- International Conference on Machine Learning (ICML)
- European Conference on Machine Learning (ECML)
- Neural Information Processing Systems (NIPS)
- Computational Learning
- International Joint Conference on Artificial Intelligence (IJCAI)
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)
- IEEE Int. Conf. on Data Mining (ICDM)

Summary

- Introductory course that covers a wide range of machine learning techniques—from basic to state-of-the-art.
- More theoretical/statistics oriented, compared to other courses I teach → might need continuous work not “to get lost”.
- You will learn about the methods you heard about: Naïve Bayes’, belief networks, regression, nearest-neighbor (kNN), decision trees, support vector machines, learning ensembles, over-fitting, regularization, dimensionality reduction & PCA, error bounds, 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
- **It's going to be fun and hard work**

Which Topics Deserve More Coverage —if we had more time?

- Graphical Models/Belief Networks (just ran out of time)
- More on Adaptive Systems
- Learning Theory
- More on Clustering and Association Analysis → covered by Data Mining Course
- More on Feature Selection, Feature Creation
- More on Prediction
- Possibly: More depth coverage of optimization techniques, neural networks, hidden Markov models, how to conduct a machine learning experiment, comparing machine learning algorithms,...