Machine Learning

By Dr Brad Morantz
Presentation to the IEEE
June 2017

Suppose You Wanted to . .

- Know if you should grant credit
- Find the best mathematician
- Invest in the stock market
- How to filter spam from email
- Detect intrusion on your network
- Implement computer vision
- Recommend a good movie to a customer
- Track the spread of influenza

Overview

- Example
- Definitions
- Why now
- Academic vs practitioner
- Supervised vs Unsupervised
- Methods
- Data Mining
- Bayesian
- Neural Networks
- Clustering
- Pros and cons
- Summary
- References

My Focus

- I am a Decision Scientist and EE
- My focus is on solving the problem
- The computer is just one of my tools
- If my only tool is a hammer, then all of the problems look like nails

Machine

- (Merriam Webster Dictionary):
  - a mechanically, electrically, or electronically operated device for performing a task
  - archaic : a constructed thing whether material or immaterial
- Wikipedia: tool containing one or more parts that uses energy to perform an intended action.
- Business Dictionary: tool containing one or more parts that uses energy to perform an intended action.

Learning

- Merriam Webster: the activity or process of gaining knowledge or skill by studying, practicing, being taught, or experiencing something
- The Conditions of Learning by Robert Gagne: A change in human disposition or capability that persists over a period of time and is not simply ascribable to processes of growth
- The Oxford Dictionary: The acquisition of knowledge or skills through study, experience, or being taught.
- Business Dictionary: Measurable and relatively permanent change in behavior through experience, instruction, or study.
- Dictionary.com:
  - 1. knowledge acquired by systematic study in any field of scholarly application.
  - 2. the act or process of acquiring knowledge or skill.
  - 3. Psychology ; the modification of behavior through practice, training, or experience.

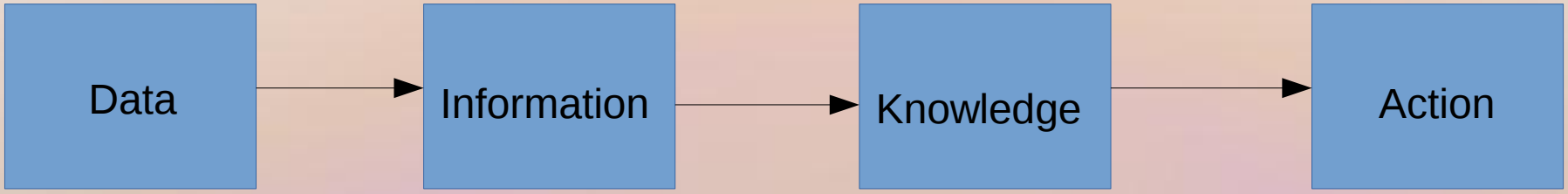
Old Saying

- A wise person learns from others' mistakes
- An average person learns from their mistakes
- A dumb person never learns
- Applying this to machines:
  - A wise machine learns from data
  - An average machine learns from its own mistakes (or when told)
  - A stupid machine does not learn

Therefore Machine Learning Is:

- An energy using tool that has the intended action of gaining knowledge or skill by studying, practicing, being taught, or experiencing something
- a mechanically, electrically, or electronically operated device for performing the task of acquiring knowledge by systematic study in any field of scholarly application
- Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. (techtarget.com)
- Machine learning explores the study and construction of algorithms that can learn from and make predictions on data (Kohavi & Provost)
- Machine learning is a method of data analysis that automates analytical model building. Using algorithms that iteratively learn from data, it allows computers to find hidden insights without being explicitly programmed where to look (SAS Institute)
- Computer system that automatically improves with experience (Tom Mitchell)

The Path



- Data are things known or assumed as facts, can be numerical, alphabetic, or symbolic
- Information is facts proved or learned about something or someone
- Knowledge is the theoretical or practical understanding of a subject
- Action is the act or process of doing something, typically to achieve a goal
  - Definitions courtesy of Dr Google

Machine Learning is **NOT**

- Machine Cognition (does not think, no cognitive function, logic, or reasoning)
- Does not always build a model
- The epitome of computer science
- The answer to solving world hunger
- All about C/C++ code

Science vs Technology

- Science is the intellectual and practical activity encompassing the systematic study of the structure and behavior of the physical and natural world through observation and experiment.

- Technology is the application of scientific knowledge for practical purposes
- We need ***both***

Definitions compliments of Dr Google

According to

- *Practitioner*
- How can we apply
- How can we increase sales
- How can we increase profit
- Reducing defects
- Increase efficiency
- About technology
- Cookbook

- *Researcher*
- What is learning
- How can we learn
- How can machine learn
- What is knowledge
- How do we gain knowledge
- Increasing science

Why Now?

- We now have lots of Data
  - Credit card
  - Frequent buyer card
  - Medical Information Bureau
  - Everything computerized collecting info
- Now have more powerful computers
  - Faster
  - Multicore & GPGPU
  - Bigger & more memory
  - ASICs & FPGA

Supervised vs Unsupervised

- Training method
- Supervised
  - Has the answer
  - It learns what patterns have that answer
- Unsupervised
  - Does NOT provide the answer
  - Learns by organizing
  - Sees what goes with what
- Semi-supervised
  - Some observations have the answer

Datamining

- Subfield of machine learning
  - Goes hunting through a data set
  - Find information, rules, models
- Unsupervised
  - Not told what is expected outcome
- Discovery process
  - Go into a data set and see what you find
  - What rules and knowledge are learned
  - Exploratory data analysis (Tukey)

Simple Example

Go to a garage sale and buy a “grab box”

You go home and sort out what you bought

- Man's hat size XXL
- Size 13W men's shoes
- A pair of extra large gloves
- A bunch of 45 RPM Beatles records

- This was from a large man about 65 y.o. who either moved or died

What is a Mountain?

A big pile of dirt

A large Amount of Data is...

- Some people would say that data on them is . . .
- Dirt
- A big pile of data is a . . .
- Mountain of data or a data warehouse
  - Coincidently they have built many data warehouses inside Iron Mountain.

Data Mining

- Is the nontrivial extraction of implicit previously unknown and potentially useful information and data. (James Buckley)
- Machine Learning employs search heuristics to uncover interesting and systematic relationships in data. (Dhar & Stein)

Data Mining

- Going into a data file and looking for patterns and relationships
- Use of historical data to find patterns and improve future decisions
- Also called KDD
  - Knowledge discovery from databases
- Seeing what things are learned from this mountain of data

First Step

- Before one eats, they wash their hands
- Before one works with data ***clean it!***
- Cleaning the data can take 75% of energy
- My abalone paper
  - Public data set
  - Did automatic cleaning program
  - Then decision tree
  - Results were better than others

Garage Sale

- You buy a box of stuff (bargain box)
- You have no idea what is in it
- You open it up
- You look through the contents
- You make note of all that you have found
- You may sort the stuff into piles based upon some similarities
- You may learn many things from the groups that you create

Think About What We Just Did

- Took a 'box' of data
- Went looking into it
- No preconceived ideas or thoughts
- Organized by what we found
- Created several piles of things in common
- Learned about what we had and maybe about the people that we bought it from

Association Rule Example

- Check out at store
  - Each check out is a record
  - Called a “basket”
- Look at what was purchased
  - Look for relationships or associated items
  - milk and cereal, beer and pizza

Grocery Example

Invoice	Milk	Cereal	Pizza	Beer	Eggs
1	X	X	X	X	
2			X	X	X
3	X	X			X
4	X	X			X
5			X	X	
6	X				X

Milk AND Cereal

- Support
  - Percent of baskets where rule is true
  - P(milk AND cereal) = 3/6 = 0.500
- Confidence
  - Percent of baskets that have cereal given milk is in the basket
  - P(cereal|milk) = ¾ = 0.75
- Knowledge gained
  - 50% improvement in ability to predict

Walmart

- IBM's single largest customer
- They study related sales
  - What people buy together (e.g. bread & jam)
  - Ratios, how much of each relative to the other
- They study where to place in store
  - Same items with different SKU's to track
- Optimize to maximize sales volume
- Optimize for JIT

Example Applications

- High risk pregnancy
  - Tons of data
  - What are common factors?
- Myocardial Infarction (MI)
  - Tons of data
  - What factors are common in non-occurrence?
- Loan application
- Many other situations

Methods (Major Classes)

- Trees
- Similarity Engine
- OLS & logistic regression
- Data visualization
- Bayesian Classification & networks
- Clustering
  - Statistical Clustering
  - Self Organizing Maps
  - K-means
  - K-nearest neighbor
- Specialized programs
  - Clementine
  - Enterprise Miner
- More

Original Project

- US Navy
  - Boilers exploding on boat
  - Expensive and difficult when not in US harbor
- ONR
  - Looked through lots of data
  - Which observation predicted boiler problems
  - Which observation did not predict boiler problems
  - Bring boats in for service before boiler explodes
- This was specific not open ended study

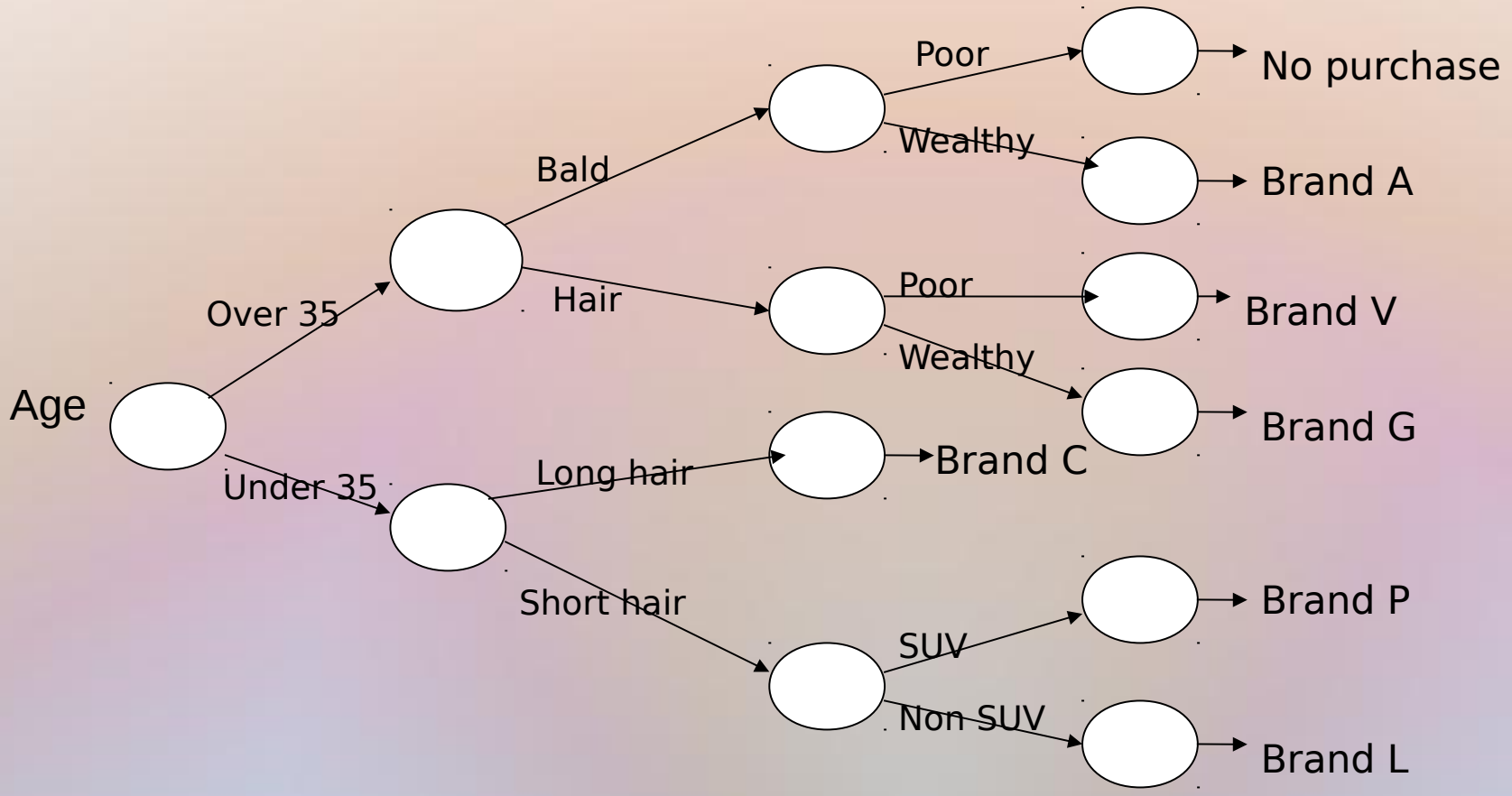
Clustering

- A dimension for each variable
- Plot each observation in N-space
- Observe locations in N-space with many points
- Distances in euclidian, manhattan, or mahlanobis
- Observe points condensing into clusters, sometimes quite tight
- Within a cluster they are similar
- To classify a new one, see which it fits into

Goals of Clustering

- Reduce variance within cluster
  - Make cluster tighter
- Maximize distance between cluster centers (centroid or medoid)
  - Clusters at least 2 sigma apart

Trees



Can start from either side and work to the other

Hierarchical (statistical) Clustering

- Can use statistics with it
- Agglomerative
  - Go from leaves to top
- Divisive
  - Top down
- My favorite
- Can extract knowledge from it by examining the clusters, sometimes need SME

Cancer Clustering

- Public data set of breast cancer
- Plotted in N-space, each variable
- Results were a few well defined clusters
- Analyzed and removed weak variables
  - Contributed more noise than information
- Clustered on new smaller data set
- Excellent results

K Nearest Neighbor

- The K nearest neighbors vote on which cluster it is best fitted for
- Some algorithms use a weighting, the closer neighbors have a stronger influence
- Used for classification or “regression”
- Supervised learning
- I am not fond of this
  - Not mathematical

K Means

- Have K clusters
- Put each observation into the cluster with the closest mean
- Partition N observations into K clusters
  - Each observation is in cluster with nearest mean
  - Need a priori knowledge (what is k?)
- Unsupervised
- Moving centroids
- Iterative process

Self Organizing Map

- A type of artificial neural network
- Unsupervised
- Competitive learning
- Similar to sensory information in human brain
- Iterative algorithm
- Need an initial approximation (a priori knowledge)
- Grossberg & Carpenter

Thomas Bayes

- 18th century English minister
- Tried to prove that G-d existed using mathematics
- Published posthumously
- Bayesian Probability was based on this

Bayesian Probability

- What if someone wanted to know the probability of having a certain disease
- If they know that the test only registers 80% correctly as having it (20% false positives)
  - P(disease | test 80% accurate)
- Could also ask probability of rain
  - P(rain | just washed my car)
- Having knowledge can improve our probability (predictive) accuracy

Bayes Probability

- P(A | B) = P(A and B) / P(B)
- P(A) is the “prior probability”
- P(A | B) is the posterior probability
  - More accurate because we have more information
  - Given some additional information, we can make a better prediction

What Does Bayes Provide

- If $P(A \mid B) = P(A)$, then B provides no new information
- A and B are independent
- If $P(A \mid B) \neq P(A)$, then B contains information and reduces entropy
- This leads to Naive Bayes Classification
  - Method pretends that only one influencing item
  - Reduces math complexity greatly

Neural Networks

- Does not build a model
- Supervised learning
- A matching system
  - Learns to match input to output
  - Universal function approximator
- No knowledge in hand
- Methods do exist to get the knowledge
  - Rule induction
  - Decision trees

Something to Think About

- Do WE have any knowledge in hand?
  - Some methods train a classifier
  - Some methods give us relationships
  - Some give us a model
  - Some give us knowledge
- Depends on what is our goal
  - Help make a decision
  - Train a robot
  - Help make a forecast
  - Help solve a problem

Required Subject Areas

- Mathematics
- Statistics
- Algorithm design
- Computer science
- Problem solving
- Thinking
- Experience
- Application subject area

Summary

- Mastering this is not a “quick learn” or cookbook
- Decide science or technology
- Understand what is knowledge & learning
- Understand the problem that needs solving
  - Abalone research example
- Learn and generate knowledge
- Act on it and solve the problem

References

- IEEE Computational Intelligence Society
  - http://www.cis.ieee.org
- KDD Nuggets http://www.kdnuggets.com
- My Website http://www.machine-cognition.com
- The library, the internet, wikipedia, etc
- Statistics & data mining class at college
- Tom Mitchell, Bruce Buchanan

Questions?