Online Machine Learning and Optimization based on Google Tools for Industry 4.0

Industry 4.0 School & Industry Night

UBC Okanagan School of Engineering

Dr Hamid Khayyam
Prof Pavel Trivailo
Prof Reza N. Jazar
Email: hamid.khayyam@rmit.edu.au

RMIT UNIVERSITY

Agenda

- Introduction Process Control: Challenges and Solutions
- Machine Learning Overview
- Industry Revolutionizing by Online Machine Learning
- Google Tools
- Carbon Nexus Industry Constructed Machine Learning (an example)
- Related our Research and Publication for Carbon Nexus
- Conclusion and Biography

Agenda

- **Introduction Process Control: Challenges and Solutions**
- Machine Learning Overview
- Industry Revolutionizing by Online Machine Learning
- Google Tools
- Carbon Nexus Industry Constructed Machine Learning (an example)
- Related our Research and Publication for Carbon Nexus
- Conclusion and Biography

Industrial Process Control: Challenges

i. Maintain certain variables within their desirable operational limits

ii. Visualized as a pyramid in complex process control

iii. Increase business value and low cost production

iv. Reduction of energy consumption

v. Limited data availability to control and optimization of the industrial process.

Process Control Our Solutions:

- ✓ Advanced control and optimization system can play an important role to improve the profitability and stability of industrial plants.
- ✓ On line control and optimization based on Cloud Computing (Google) is very useful for control of complex process
- ✓ Machine learning in which (Limited)data becomes available in a sequential order and is used to update our best predictor for future data at each step,
- ✓ Online learning is computationally infeasible to train over the entire dataset, requiring the need of out-of-core algorithms.
- ✓On time, machine learning algorithms based on Google tools, has the potential to bring greater predictive accuracy to every phase of production.

Agenda

- Introduction Process Control: Challenges and Solutions
- **Machine Learning Overview**
- Industry Revolutionizing by Online Machine Learning
- Google Tools
- Carbon Nexus Industry Constructed Machine Learning (an example)
- Related our Research and Publication for Carbon Nexus
- Conclusion and Biography

Online machine learning is ….

- Machine learning in which data becomes available in a sequential order and is used to update our best predictor for future data at each step,

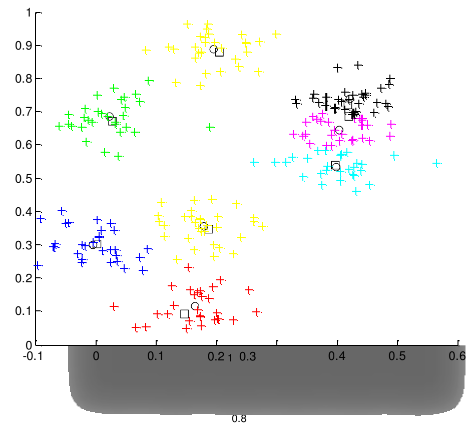
- Online learning is computationally infeasible to train over the entire dataset, requiring the need of out-of-core algorithms.

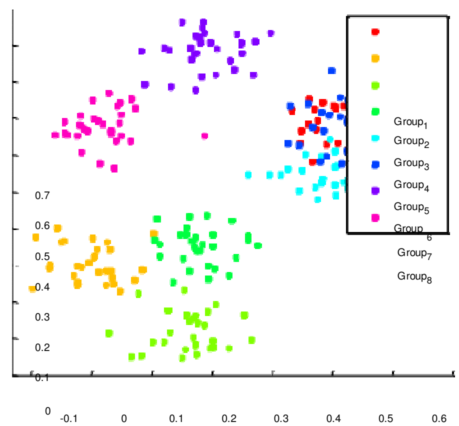
- Online learning machine learning's core technologies align well with the complex problems in industry.

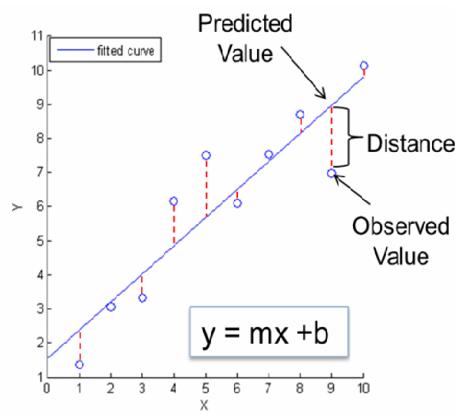
- On time, machine learning algorithms have the potential to bring greater predictive accuracy to every phase of production.

Machine Learning ….

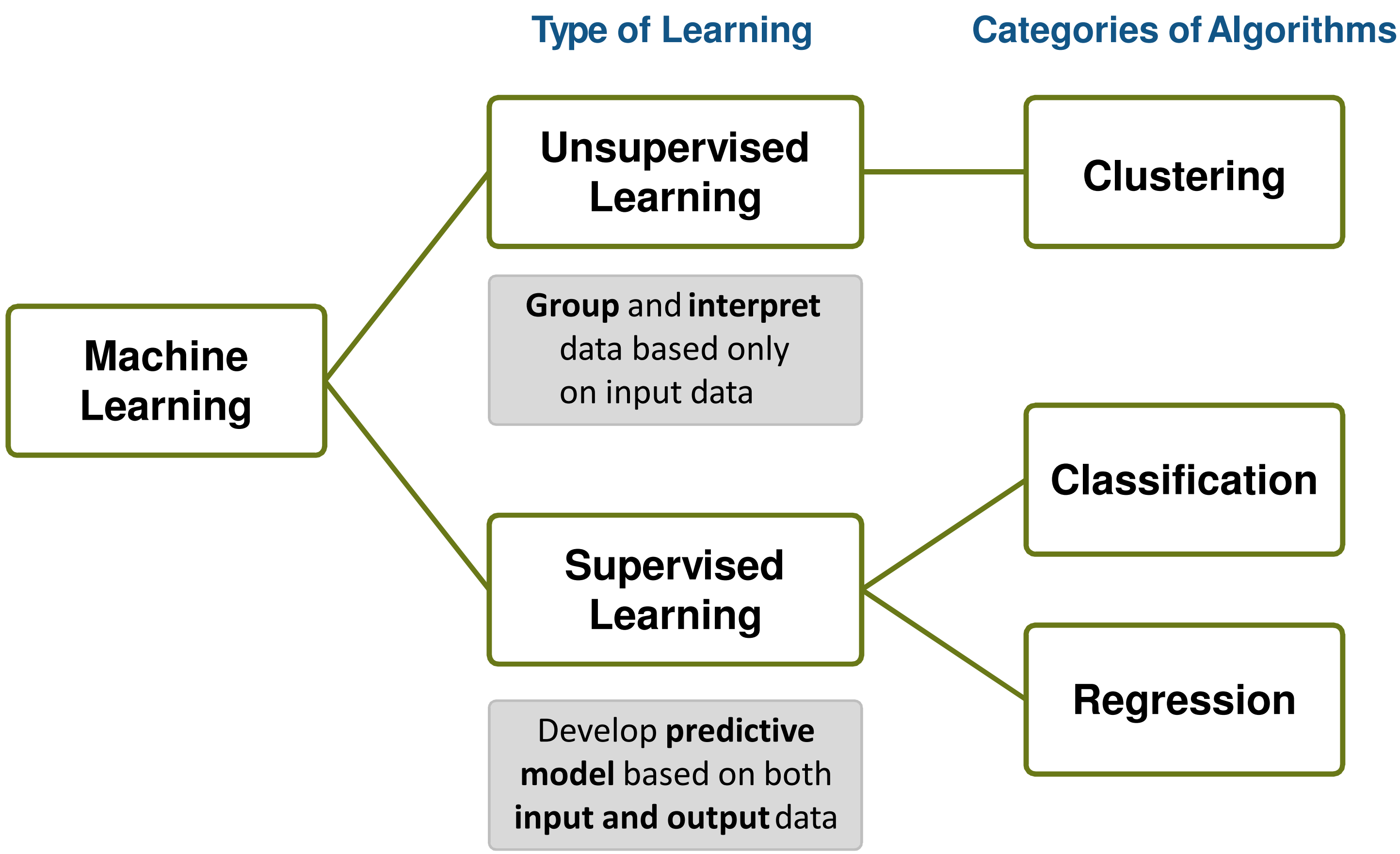
- Machine Learning Overview
- Machine Learning
  - Unsupervised Learning
    - Clustering
  - Supervised Learning
    - Classification
    - Regression
- Learn More



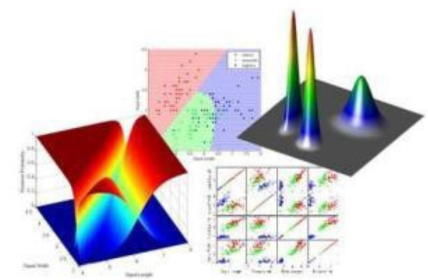


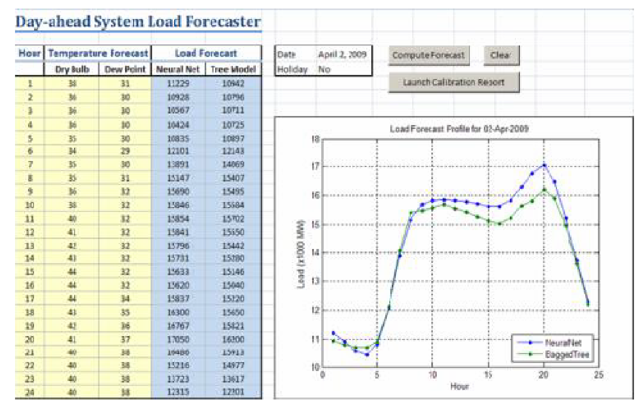


Machine Learning Overview
Types of Learning, Categories of Algorithms….

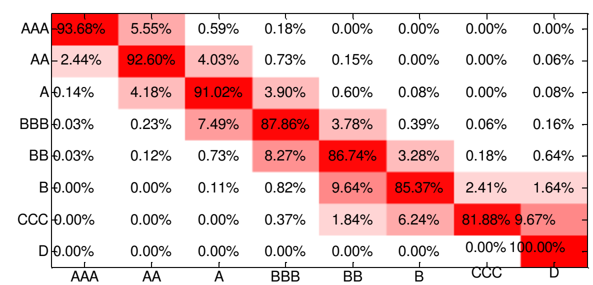


Machine Learning
When and where it is used?



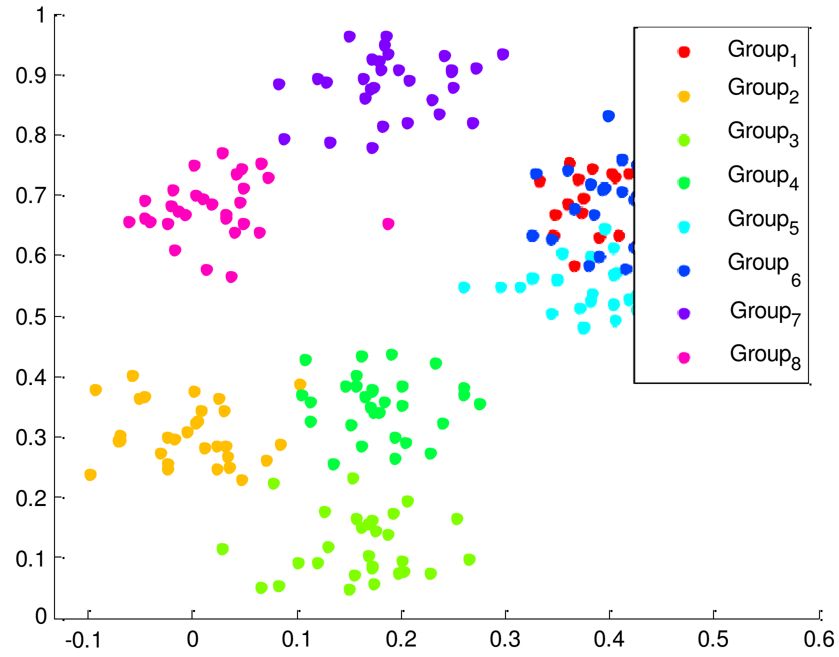


- When to use it
  - Predict a future outcome based on
    - Historical data (many variables)
    - Specific patterns
  - Define a System that is
    - Based on inputs and outputs from the system
    - complex to define using governing equations
      (e.g., black-box modeling)
- Examples
  - Pattern recognition *(speech, images)*
  - Financial algorithms *(credit scoring, algo trading)*
  - Energy forecasting *(load,consumption, price)*
  - Biology *(tumor detection, drug discovery)*



Basic Concepts in Machine Learning

- Start with an initial set of data
- "**Learn**" from this data
  - "**Train**" your algorithm with this data
- Use the resulting **model** to **predict** outcomes for **new** data sets

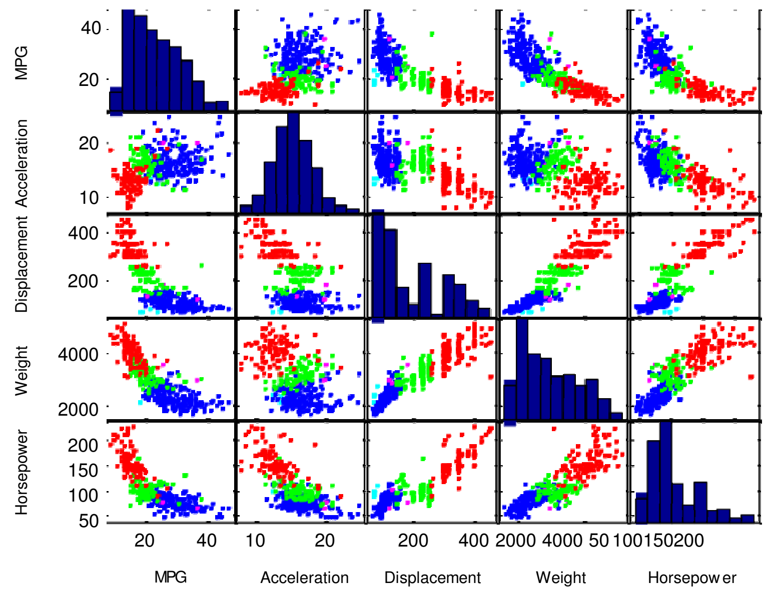


Machine Learning Process

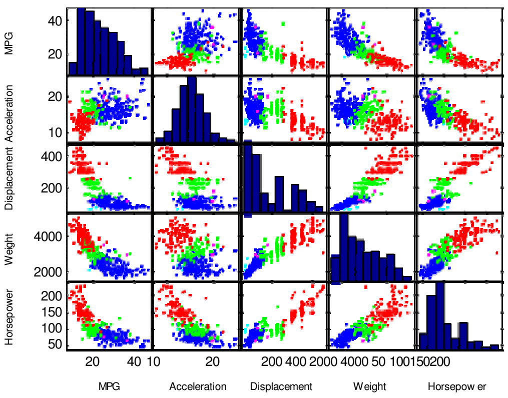
Exploration → Modeling → Evaluation → Deployment

Exploratory Data Analysis

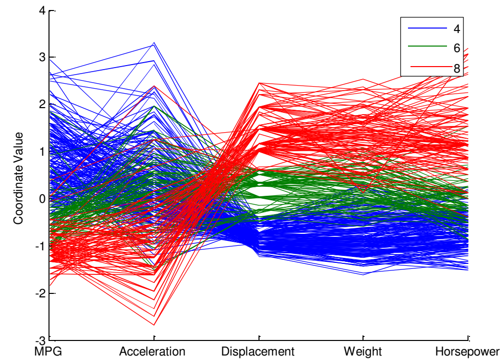
- Gain insight from visual examination
  - Identify trends and interactions
  - Detect patterns
  - Remove outliers
  - Shrink data
  - Select and pare predictors
  - Feature transformation



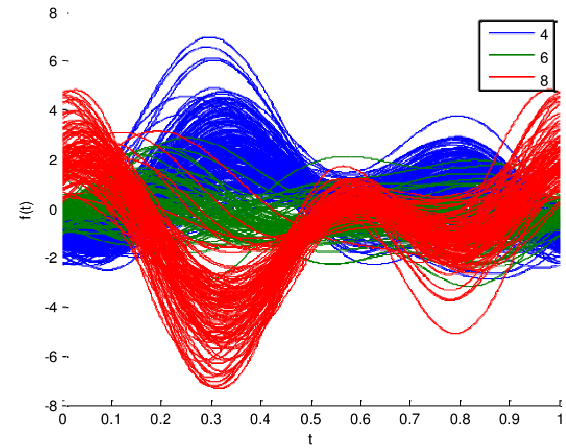
Data Exploration
Interactions Between Variables



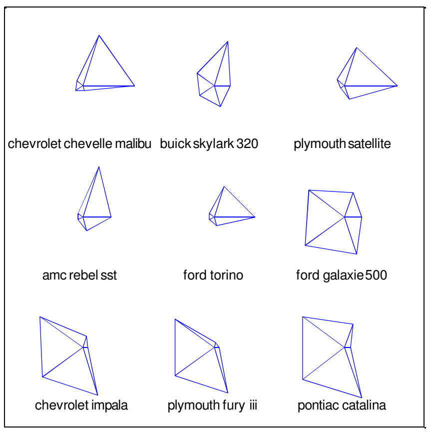
Plot Matrix by Group



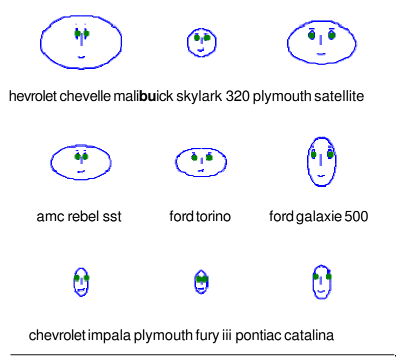
Parallel Coordinates Plot



Andrews' Plot

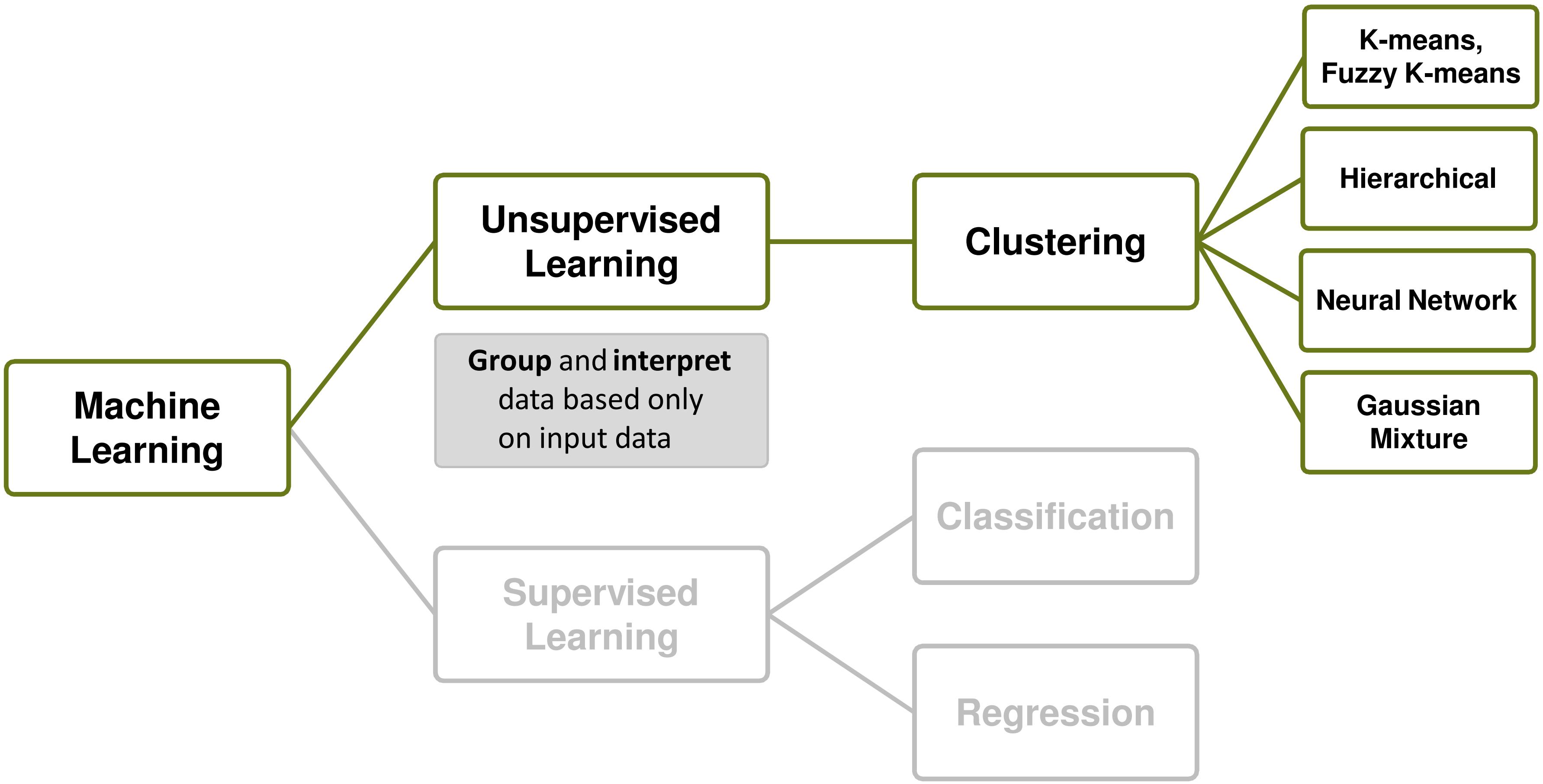


Glyph Plot



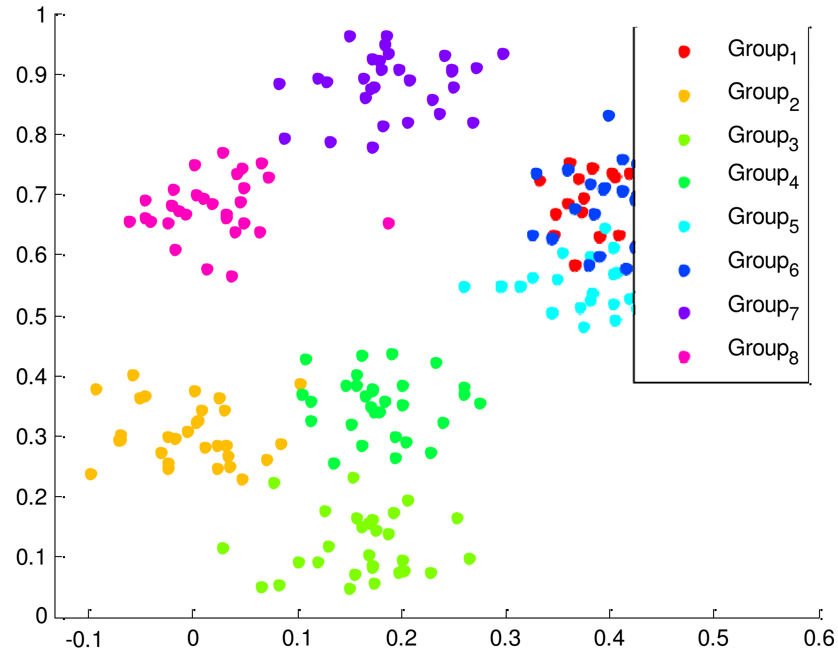
Chernoff Faces

Unsupervised Learning
Clustering



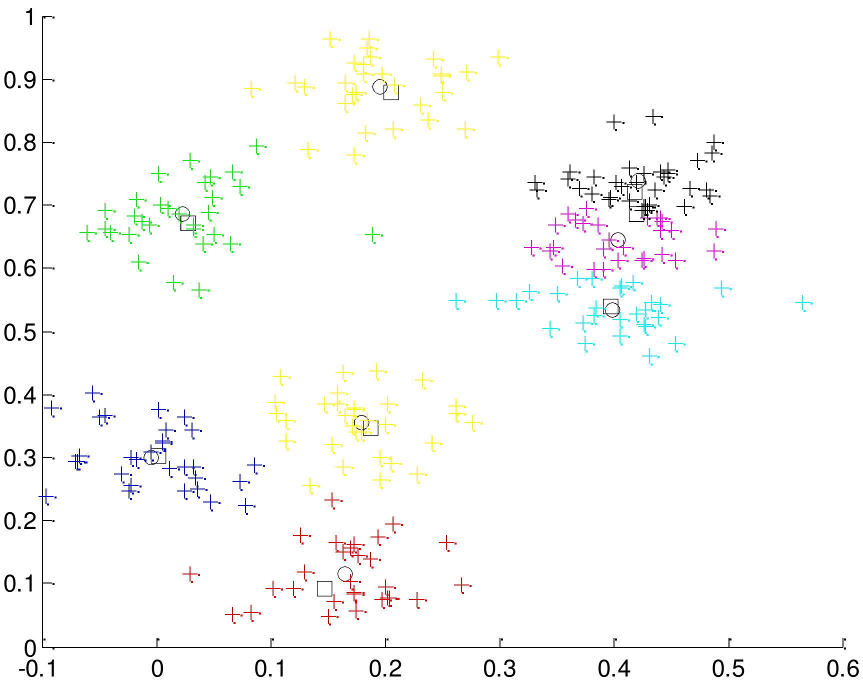
Dataset We’ll Be Using

- Cloud of randomly generated points
  - Each cluster center is randomly chosen inside specified bounds
  - Each cluster contains the specified number of points per cluster
  - Each cluster point is sampled from a Gaussian distribution
  - Multi-dimensional dataset



Clustering
Overview

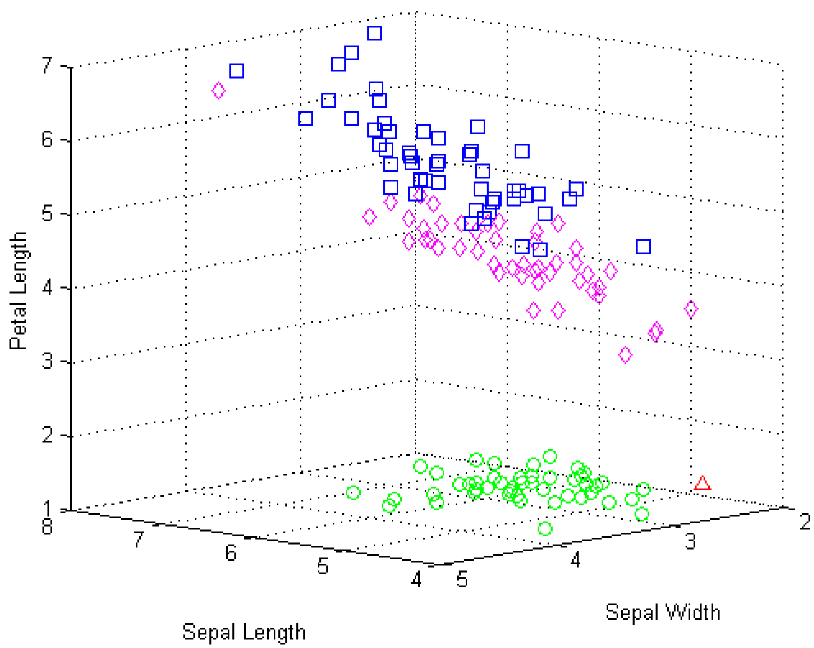
- What is clustering?
  – Segment data into groups, based on data similarity
- Why use clustering?
  – Identify outliers
  – Resulting groups may be the matter of interest



- How is clustering done?
  – Can be achieved by various algorithms
  – It is an iterative process *(involving trial and error)*

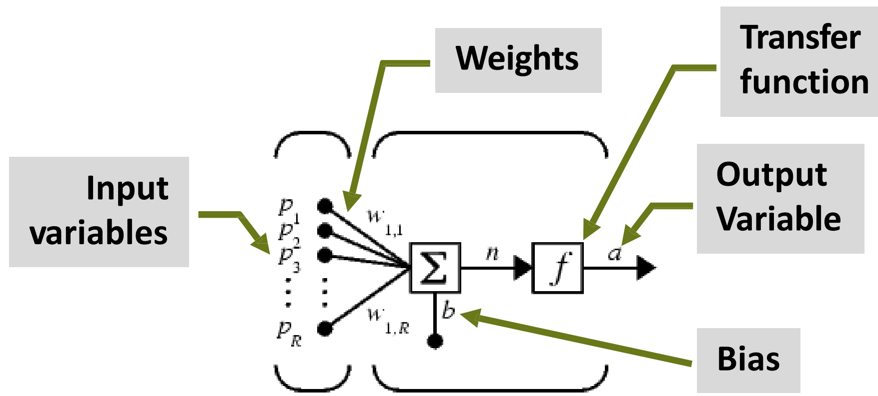
Clustering: K – Means Clustering

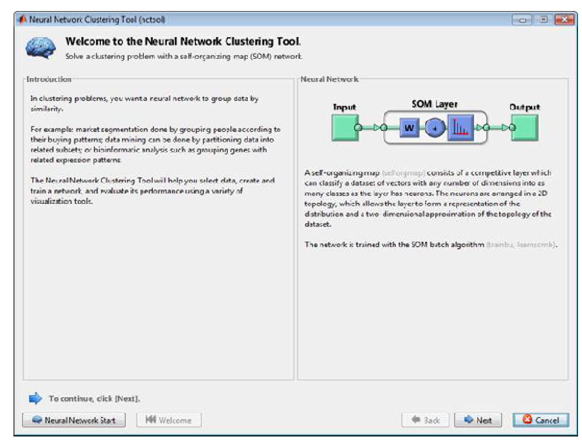
- K-means is a partitioning method
- Partitions data into ***k*** mutually exclusive clusters
- Each cluster has a centroid (or center)
  – Sum of distances from all objects to the center is minimized



Clustering: Neural Networks

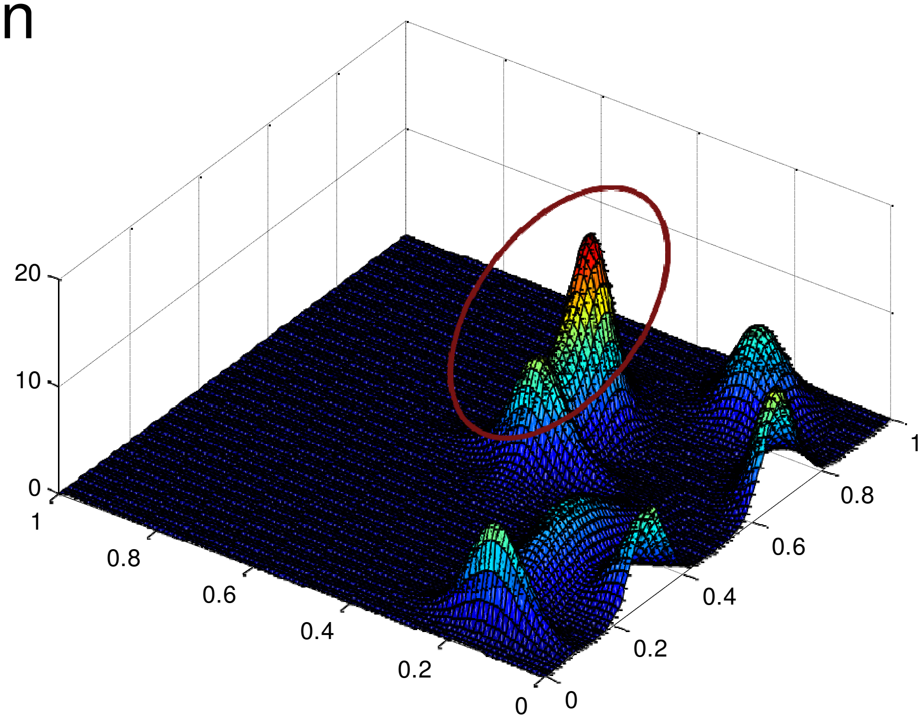
- Networks are comprised of one or more layers
- Outputs computed by applying a nonlinear transfer function with weighted sum of inputs
- Trained by letting the network continually adjust itself to new inputs *(determines weights)*
- Interactive apps for easily creating and training networks
- Multi-layered networks created by cascading (*provide better accuracy)*
- Example architectures for clustering:
  - Self-organizing maps
  - Competitive layers





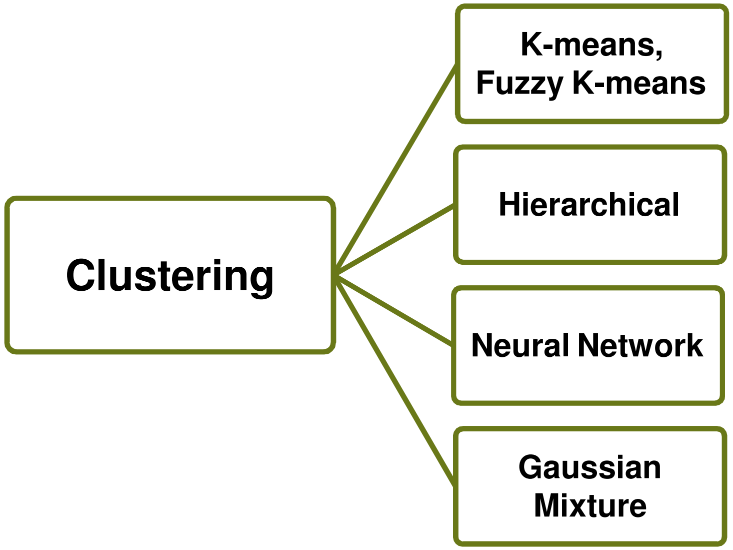
Clustering : Gaussian Mixture Models

- Good when clusters have different sizes and are correlated
- Assume that data is drawn from a fixed number ***K*** of normal distributions



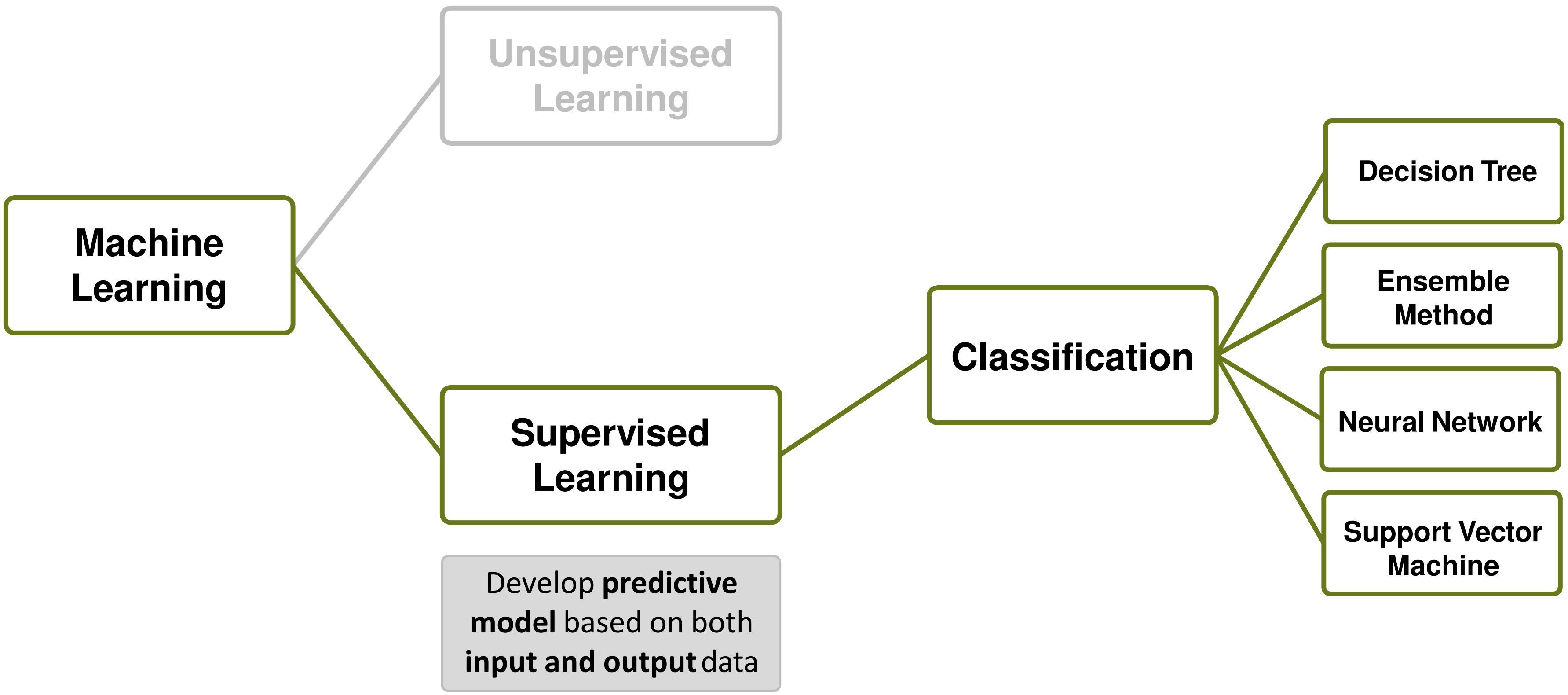
Cluster Analysis
Summary

- Segments data into groups, based on data similarity
- No method is perfect
  (depends on data)
- Process is iterative;
  explore different algorithms
- Beware of local minima
  (global optimization can help)



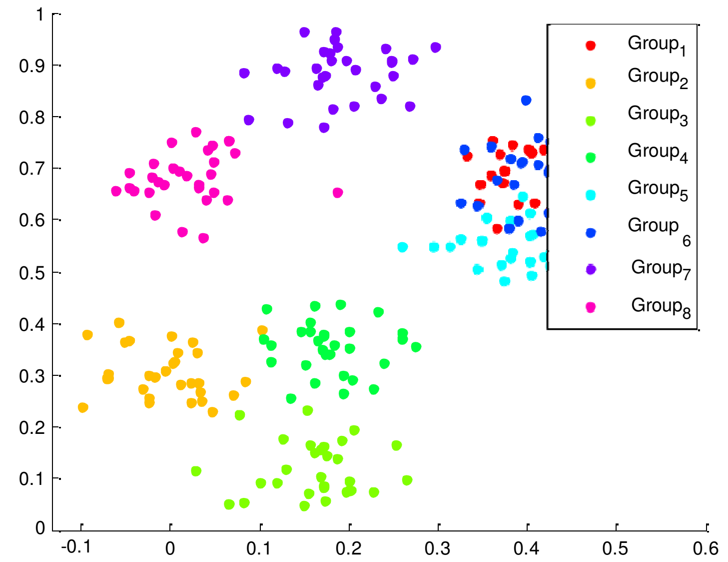
Model Development Process

Supervised Learning
Classification for Predictive Modelling



Classification
Overview

- What is classification?
  – Predicting the best group for each point
  – “Learns” from labeled observations
  – Uses input features

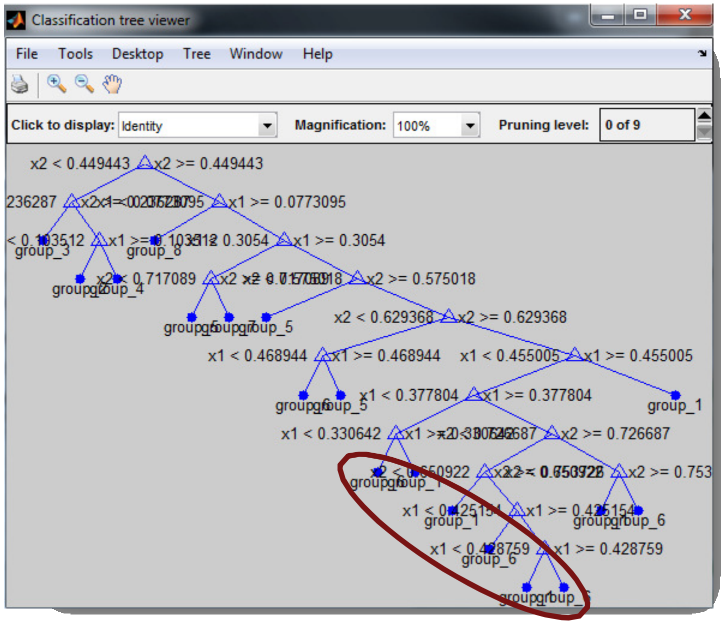


- Why use classification?
  – Accurately group data never seen before

- How is classification done?
  – Can use several algorithms to build a predictive model
  – Good training data is critical

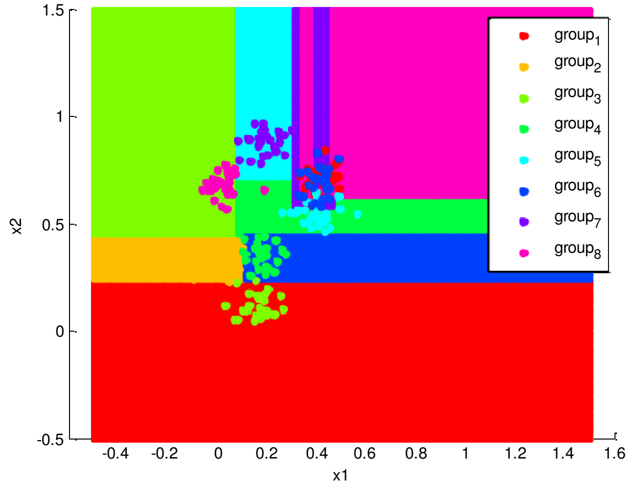
Classification - Decision Trees

- Builds a tree from training data
  - Model is a tree where each node is a logical test on a predictor
- Traverse tree by comparing features with threshold values
- The “leaf” of the tree specifies the group



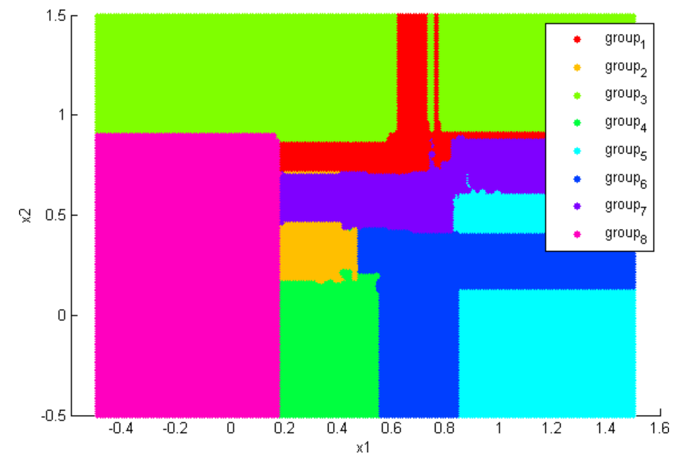
Classification - Ensemble Learners
Overview

- Decision trees are “weak” learners
  - Good to classify data used to train
  - Often not very good with new data
  - Note rectangular groups



- What are ensemble learners?
  - Combine many decision trees to create a “strong” learner
  - Uses “bootstrapped aggregation”

- Why use ensemble methods?
  - Classifier has better predictive power
  - Note improvement in cluster shapes

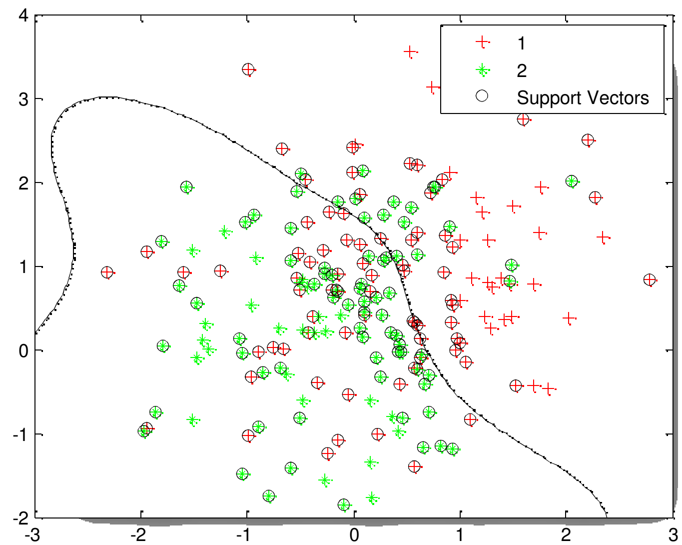


Classification - Support Vector Machines
Overview

- Good for modeling with complex boundaries between groups
  - Can be very accurate
  - No restrictions on the predictors

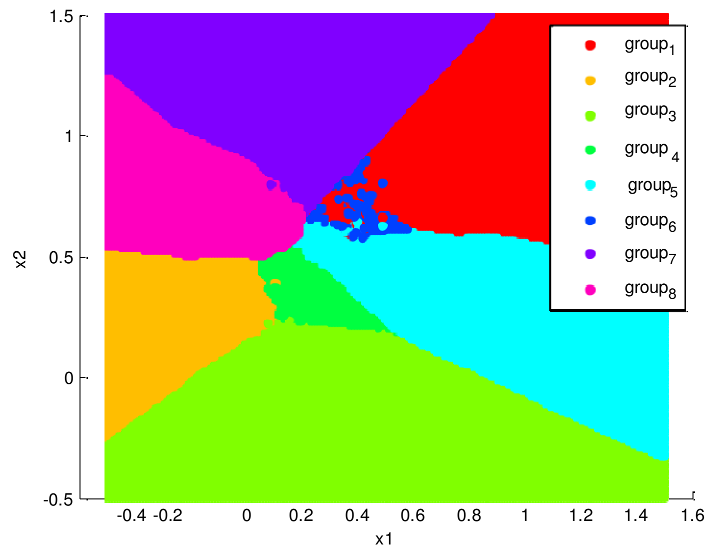
- What does it do?
  - Uses non-linear “kernel” to calculate the boundaries
  - Can be computationally intensive

- Version in Statistics Toolbox only classifies into two groups



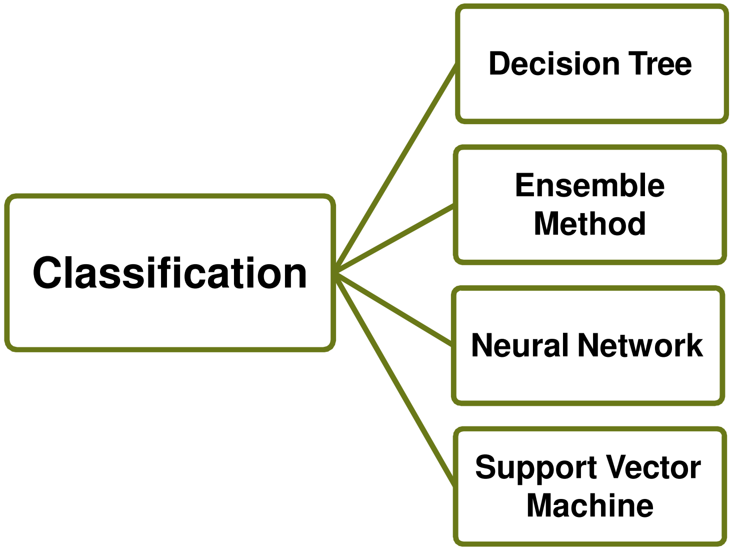
K-Nearest Neighbor Classification

- One of the simplest classifiers
- Takes the ***K*** nearest points from the training set, and chooses the majority class of those ***K*** points
- No training phase – all the work is done during the application of the model

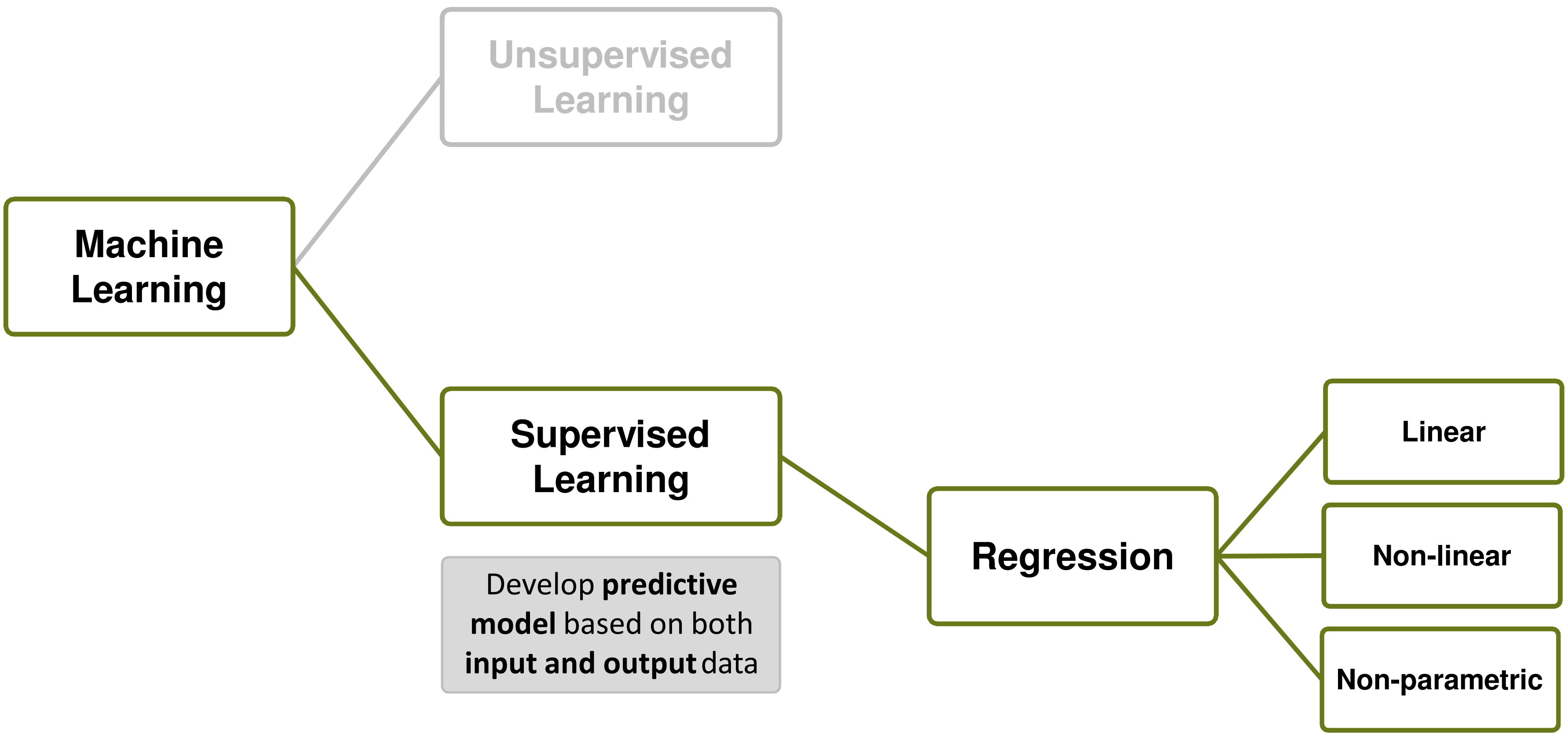


Classification
Summary

- No absolute best method
- Simple does not mean inefficient
- Watch for overfitting
  - Decision trees and neural networks may overfit the noise
  - Use ensemble learning and cross-validation
- Parallelize for speedup

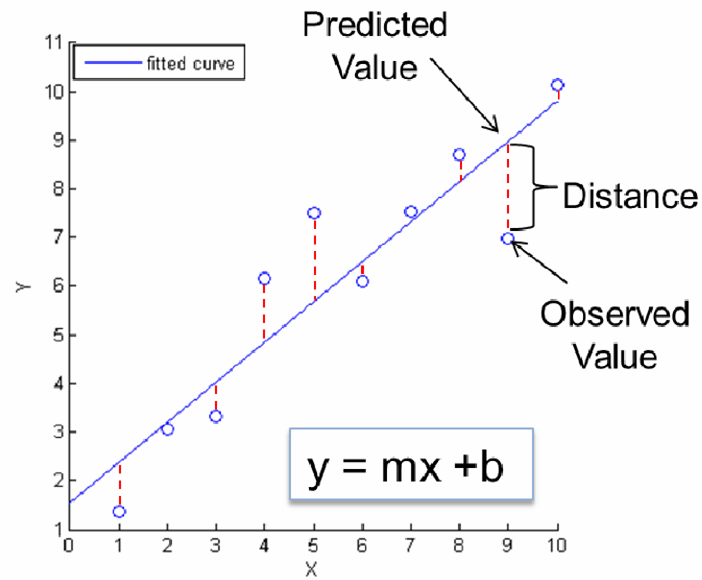


Supervised Learning
Regression for Predictive Modelling



Regression

- Why use regression?
  - Predict the continuous response for new observations
- Type of predictive modeling
  - Specify a model that describes Y as a function of X
  - Estimate coefficients that minimize the difference between predicted and actual



- You can apply techniques from earlier sections with regression as well

Linear Regression

- Y is a *linear* function of the regression coefficients
- Common examples:

Straight line	$Y = B_0 + B_1 X_1$
Plane	$Y = B_0 + B_1 X_1 + B_2 X_2$
Polynomial	$Y = B_0 + B_1 X_{1_3} + B_2 X_{1_2} + B_3 X_1$
Polynomial with cross terms	$Y = B_0 + B_1 X_1^2 + B_2 (X_1 * X_2) + B_3 X_2^2$

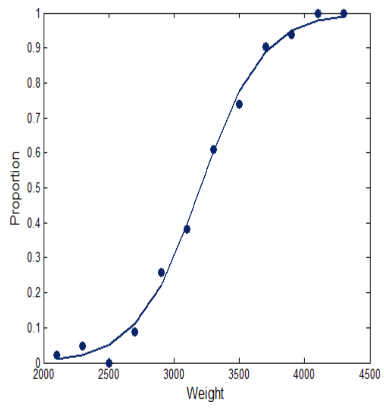
Nonlinear Regression

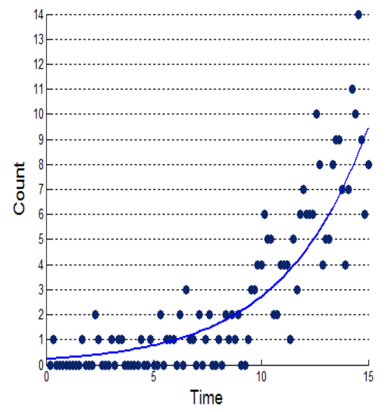
- Y is a *nonlinear* function of the regression coefficients
- Syntax for formulas:

Fourier Series	`y ~ b0 + b1*cos(x*b3) + b4*sin(x*b3)`
Exponential Growth	`@(b,t)(b(1)*exp(b(2)*t)`
Logistic Growth	`@(b,t)(1/(b(1) + exp(-b(2)*x)))`

Generalized Linear Models

- Extends the linear model
  - Define relationship between model and response variable
  - Model error distributions other than normal
- Logistic regression
  - Response variable is binary (true / false)
  - Results are typically expressed as an odd's ratio
- Poisson regression
  - Model count data (non-negative integers)
  - Response variable comes from a Poisson distribution





Agenda

- ❑ Introduction Process Control Challenges and Solutions
- ❑ Machine Learning Overview
- ❑ **Industry Revolutionizing by Online Machine Learning**
- ❑ Google Tools
- ❑ Carbon Nexus Industry Constructed Machine Learning (an example)
- ❑ Related our Research and Publication for Carbon Nexus
- ❑ Conclusion and Biography

Industry Revolutionizing by Online Machine Learning:

1- Increasing production capacity while lowering material consumption.

2- Providing more relevant data so finance, operations, and supply chain teams can better manage factory and demand-side constraints.

3- Enabling condition monitoring processes that provide manufacturers with the scale to manage.

4- Machine learning is revolutionizing relationship intelligence and Salesforce is quickly emerging as the leader.

5- Revolutionizing product and service quality with machine learning algorithms that determine which factors most and least impact quality company-wide.

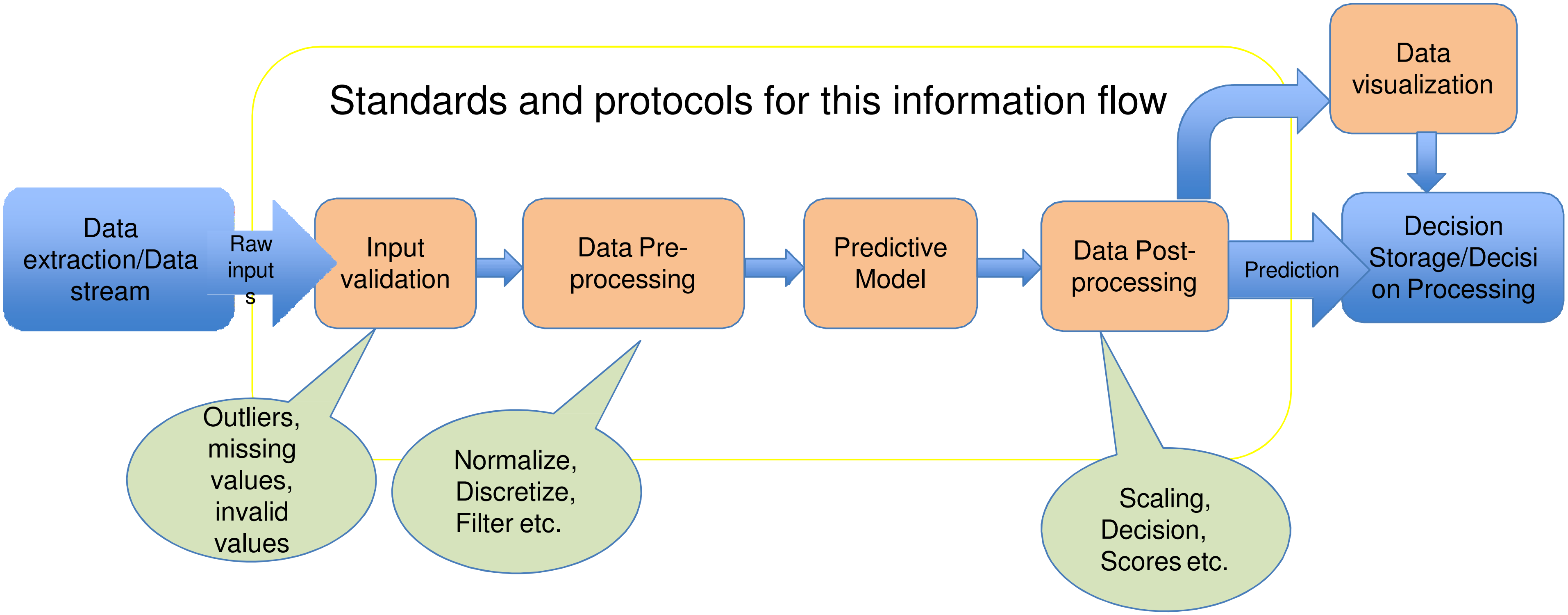
Online Machine Learning for industry 4.0

- ❑ Machine learning is revolutionizing relationship intelligence and Salesforce is quickly emerging as the leader.
- ❑ Our objective is to deliver measurement science, standards and protocols, and tools needed to predict, assess, optimize, and control the performance of manufacturing systems.

Major Our Projects:

- ➢ Machine learning methodology and associated tools to predict, assess, and optimize the operational performance
- ➢ Limited experimental data, methods , tools, data analytics and associated methods and tools to enable adaptive system

Online Predictive Analytics Workflow



Agenda

- ❑ Introduction Process Control Challenges and Solutions
- ❑ Machine Learning Overview
- ❑ Industry Revolutionizing by Online Machine Learning
- ❑ **Google Tools**
- ❑ Carbon Nexus Industry Constructed Machine Learning (an example)
- ❑ Related our Research and Publication for Carbon Nexus
- ❑ Conclusion and Biography

Google Tools (Our Solution)

- Google specializes in Internet-related services and products, which include online advertising technologies, search engine, future trends in cloud computing, software, and hardware.

- Google popular services designed are:
  Google Docs, Sheets, Slides, Email (Gmail/Inbox), Scheduling and time management (Google Calendar), Cloud storage (Google Drive), social networking (Google+), Instant messaging and video chat (Google Allo/Duo/Hangouts), Mapping and turn-by-turn navigation (Google Maps/Waze/Earth/Street View), Video sharing (YouTube), note-taking (Google Keep), Photo organizing and editing (Google Photos), Android mobile operating system, Google Chrome web browser, and Chrome OS.

- Google optimization programming is one of the latest Google services for computing the best solution to a problem modelled as a set of model and relationships.

Google Tools: Optimization Programming

Google Optimization suite contains:

- A constraint programming solver.
- A simple and unified interface to several linear programming and mixed integer programming solvers, including CBC, CLP, GLOP, GLPK, Gurobi, CPLEX, and SCIP.
- Graph algorithms (shortest paths, min cost flow, max flow, linear sum assignment).
- Algorithms for the Traveling Salesman Problem and Vehicle Routing Problem.
- Bin packing and knapsack algorithms.
- Google created OR-Tools in C++, but you can also use it with Python, C#, or Java.

Google Tools: Optimization Programming

- Google provides three ways to solve optimization problems:
  - the open-source library Glop,
  - the Linear Optimization add-on for Google Sheets, and
  - the Linear Optimization Service in Google Apps Script.

- **Glop** is Google's in-house linear solver, available as open source. You can access Glop through the OR-Tools linear solver wrapper, which is a wrapper for Glop, as well as several other third-party linear optimization solvers.
- **The Linear Optimization add-on for Google Sheets** lets you solve linear optimization problems by entering variables and constraints in a spreadsheet. Under the hood, it uses Apps Script's Linear Optimization Service.
- **The Linear Optimization Service in Google Apps Script** lets developers make function calls to solve linear optimization problems. It relies on Glop for pure linear-optimization problems where all variables can take on real values. If any variables are constrained to be integers, the service uses SCIP, from Zuse-Institut Berlin.

Google Tools: Optimization Programming

- Some of the illustrations of the Google Optimization Tools:

  - At Google, engineers work on plenty of optimization problems. One example is their YouTube **video stabilization system**, which uses linear optimization to eliminate the shakiness of handheld cameras.

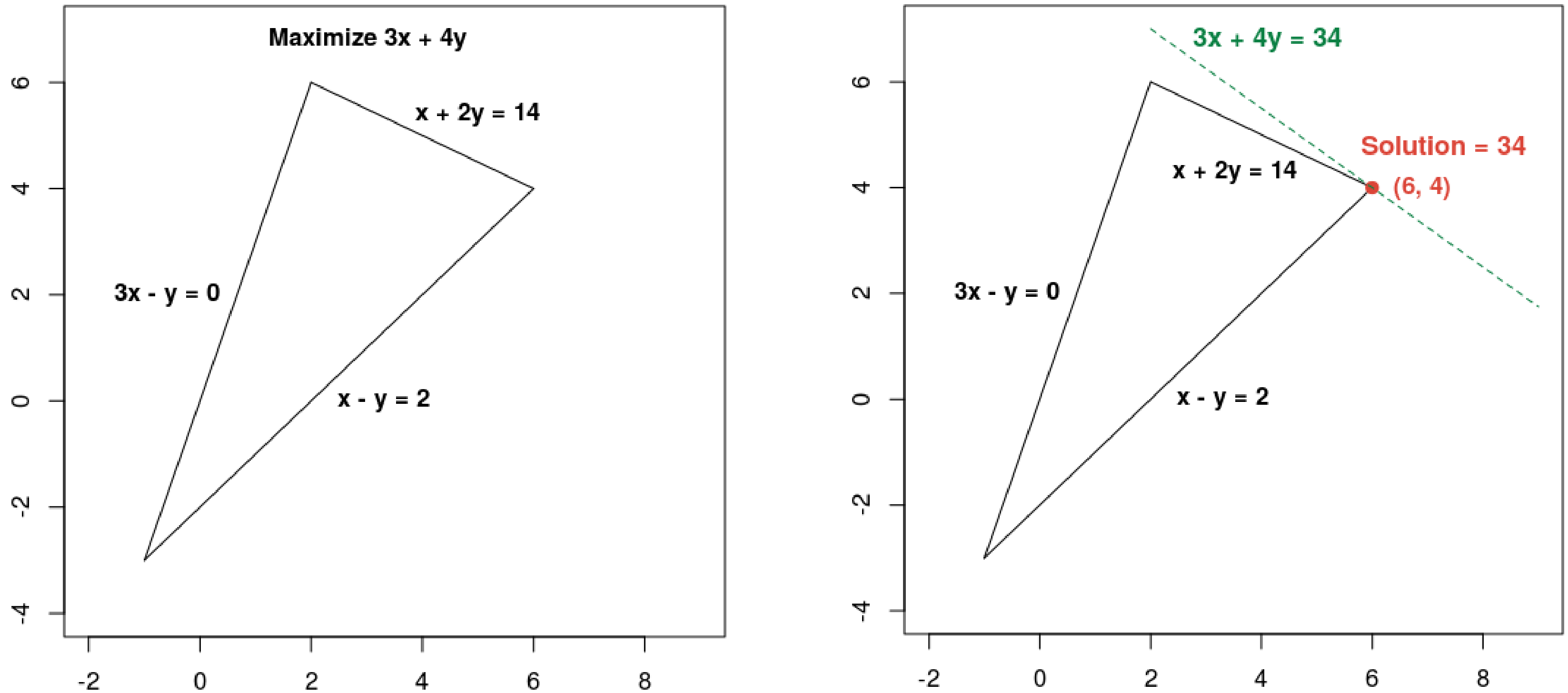
  - A more lighthearted example is in the Google Docs **Sudoku** add-on, which instantaneously generates and solves Sudoku puzzles inside a Google Sheet, using the SCIP mixed integer programming solver to compute the solution.

  - As another illustration, Google presents Glop solution to the classical **Stigler diet problem**. Stigler posed his problem as follows: given nine nutrients (calories, protein, Vitamin C, and so on) and 77 candidate foods, find the foods that could sustain soldiers at minimum cost. In 1947, Jack Laderman used Simplex, nine calculator-wielding clerks, and 120 person-days to arrive at the optimal solution. Glop's Simplex implementation solves the problem in 300 milliseconds.

Google Tools: Linear Optimization example

- Some of the illustrations of the Google Optimization Tools:

  - Graphical illustration of the use Google Optimization Tools is illustrated with the following problem:



Courtesy: https://developers.google.com/optimization/introduction/using

Google Tools: Optimization Programming

- Reasons to consider Google Optimization technology as a future trend:
  - It is constantly evolving,
  - It is a cloud based technology,
  - Open source and free,
  - Actively maintained,
  - Documented,
  - Portable & Efficient,
  - User-friendly,
  - Well tested,
  - It is a service naturally integrated with other services, for example, real-time internet search, offering powerful information support,
  - Google Application Scripting enables application of the technology to variety of particular tasks and enables adjustment to their specific requirements,
  - Has prospects for extension to the non-linear programming tasks, etc.
  - **Frontline Systems** has released a free Solver add-on for Google Sheets that solves not only linear optimization problems, but **nonlinear** ones as well.

Google Tools: Machine Learning

- Google's machine-learning software has learned to replicate itself.
- Google has announced that AutoML has beaten the human AI engineers at their own game by building machine-learning software that's more efficient and powerful than the best human designed systems.
- An AutoML system recently broke a record for categorizing images by their content, scoring 82 percent.
- While that's a relatively simple task, AutoML also beat the human-built system at a more complex task integral to autonomous robots and augmented reality: marking the location of multiple objects in an image. For that task, AutoML scored 43 percent versus the human-built system's 39 percent.
- Google is aiming to hone AutoML until it can function well enough for programmers to use it for practical applications. If they succeed in this, AutoML is likely to have an impact far beyond the walls of Google.

Machine Learning now in Google Sheets

- TBigML add-on provides an easy access to the models and clusters under user's account in BigML within the Google Sheet: https://bigml.com/tools/bigml-gas

- Using them, the blank cells in user's datasheet can be filled with the predictions that the model can extract from the known values in each data row.

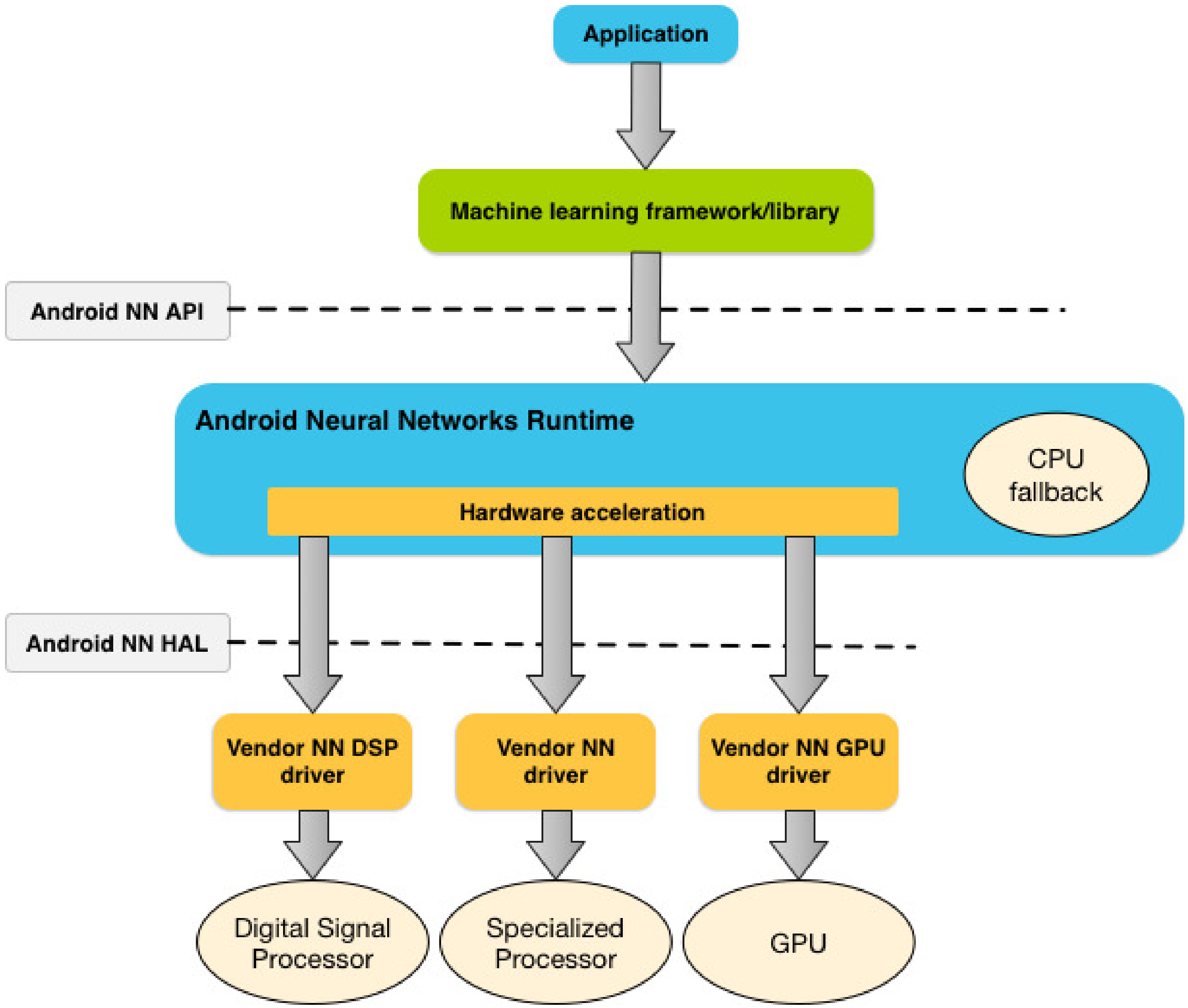
Google's Neural Networks

- In Oct 2017 Google introduced **Neural Networks API** in developer preview of Android 8.1

- New Neural Networks API brings hardware-accelerated inference to the phone for quickly executing previously trained machine learning models. Bringing these calculations to the edge can bring a lot of utility to the end user by reducing latency and load on the network, while also keeping more sensitive data on-device.

- This can come in handy when it comes to allowing the apps on user's phone to do things like classify images or learn from how your habits predict behavior. Google said they designed the Neural Networks API as a "foundational layer" for frameworks like TensorFlow Lite, Caffe2 and others.

Google Tools: Neural Networks (Cont'd)



Courtesy: https://techcrunch.com/2017/10/25/google-introduces-neural-networks-api-in-developer-preview-of-android-8-1/

Agenda

- Introduction Process Control Challenges and Solutions
- Machine Learning Overview
- Industry Revolutionizing by Online Machine Learning
- Google Tools
- **Carbon Nexus Industry Constructed Machine Learning (an example)**
- Related our Research and Publication for Carbon Nexus
- Conclusion and Biography

Carbon Nexus Industry Constructed Machine Learning

A video from Carbon Nexus

https://www.youtube.com/watch?v=tEMLk7mXOi4

Dr Hamid Khayyam (PhD, SMIEEE)
Research Fellow at Carbon Nexus (2013-2016)
Senior Lecturer at RMIT University (2017-now)

Agenda

- Introduction Process Control Challenges and Solutions
- Machine Learning Overview
- Industry Revolutionizing by Online Machine Learning
- Google Optimization Tools
- Carbon Nexus Industry Constructed Machine Learning (an example)
- **Related our Research and Publication for Carbon Nexus**
- Conclusion and Biography

Some of Related our Research and Publications

- [1] **H Khayyam**, M Naebe, O Zabihi, R Zamani, S Atkiss, B Fox, (2015)"*Dynamic Prediction Models and Optimization of polyacrylonitrile (PAN) Stabilization Processes for Production of Carbon Fiber*", IEEE Transaction on Industrial Informatics 11 (4), 887-896.
- [2] **H Khayyam**, M Naebe, A Bab-Hadiashar, F Jamshidi, Q Li, S Atkiss,… (2015),"*Stochastic optimization models for energy management in carbonization process of carbon fiber production*" Applied energy 158, 643-655.
- [3] **H. Khayyam**, G. Golkarnarenji, R. N. Jazar, (2018) "*Limited data modelling approaches for engineering applications*" Nonlinear Approaches in Engineering Applications Pp 345-379, Publisher Springer, Cham.
- [4] **H Khayyam**, SM Fakhrhoseini, JS Church, AS Milani, A Bab-Hadiashar, ... (2017) "*Predictive modelling and optimization of carbon fiber mechanical properties through high temperature furnace*" Applied Thermal Engineering 125, 1539-1554.
- [5] G Golkarnarenji, M Naebe, JS Church, K Badii, A Bab-Hadiashar, **H. Khayyam** (2017) "*Development of a predictive model for study of skin-core phenomenon in stabilization process of PAN precursor*", Journal of Industrial and Engineering Chemistry 49, 46-60
- [6] G Golkarnarenji, M Naebe, K Badii, AS Milani, RN Jazar, **H Khayyam** (2018) "*Support vector regression modelling and optimization of energy consumption in carbon fiber production line*",Computers & Chemical Engineering 109, 276-288
- [7] K Badii, JS Church, G Golkarnarenji, M Naebe, **H Khayyam** (2016) "*Chemical structure based prediction of PAN and oxidized PAN fiber density through a non-linear mathematical model*". Polymer Degradation and Stability 131, 53-61

Agenda

- ❑ Introduction Process Control Challenges and Solutions
- ❑ Machine Learning Overview
- ❑ Industry Revolutionizing by Online Machine Learning
- ❑ Google Tools
- ❑ Carbon Nexus Industry Constructed Machine Learning (an example)
- ❑ Related our Research and Publication for Carbon Nexus
- ❑ **Conclusion and Biography**

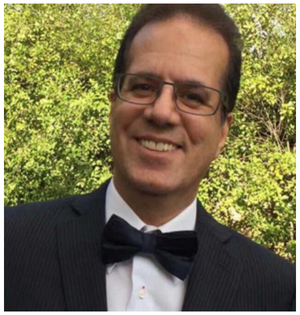
Conclusion

- ✓ Every industry has the potential to integrate machine learning into their operations and become more competitive by gaining predictive insights into production.
- ✓ Online machine learning serves to improve efficiency in the overall operation. By quickly returning predictions on the impact of minor changes, decision-makers can understand the effects before they implement something new.
- ✓ Google latest developments are opening new opportunities for solving wide range of engineering problems and process control, using machine larding.
- ✓ Google application scripts, providing easy ways to automate tasks across Google platforms, is an essential element in this cluster of new cloud-based machine learning and optimization technologies.

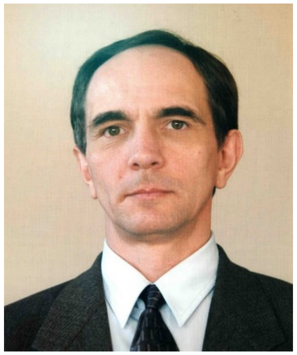
Biography of Presenters

Prof Reza Jazar (reza.Jazar@rmit.edu.au)
Professor of Mechanical Engineering (RMIT University)
Published more than 200 peer journal and conference paper and ten books.
Google scholar h-index 22 with 2700 citations.

Research interests: Nonlinear Dynamics and Vibrations, Robotics, MEMS



Prof Pavel Trivailo (trivailo@rmit.edu.au)
Professor of Aerospace Engineering (RMIT University)
Published more than 200 peer journal and conference paper and four book chapters.
Google scholar h-index 14 with 550 citations.
Research interests: Dynamics, modelling, control and optimization of complex systems, VR.



Dr Hamid Khayyam (hamid.Khayyam@rmit.edu.au)
Senior Lecturer (RMIT University)
More than 10 years worked in industry as project manager and engineering manager.
Published more than 70 peer reviewed journal and conference paper and two book chapters.
Google scholar h-index 18 with 890 citations.

Research interests: Modelling, control and optimization of complex systems.

References

- Alex Smola and S.V.N. Vishwanathan "Introduction to Machine Learning" Cambridge University Press, 2008.
- Machine Learning with MATLAB by U. M. Sundar MathWorks India 2013.
- The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc 2016.
- S. B. Kotsiantis "Supervised Machine Learning: A Review of Classification Techniques" Informatica vol 31 (2007) pp 249-268.
- Google; https://developers.google.com/optimization/lp ; https://research.googleblog.com/2014/09/sudoku-linear-optimization-and-ten-cent.html ; https://research.googleblog.com/2014/09/sudoku-linear-optimization-and-ten-cent.html
- 10 Ways Machine Learning Is Revolutionizing Manufacturing https://www.forbes.com/sites/louiscolumbus/2016/06/26/10-ways-machine-learning-is-revolutionizing-manufacturing/#22dfdb4a28c2