AI TECH MCR-2

AI User Group

10th Sep 2018

#AIUserGroup @aitechUG

Guest speaker: Fahd Rafi, Microsoft

Host: Sherin Mathew

Partnered with

Sponsored by

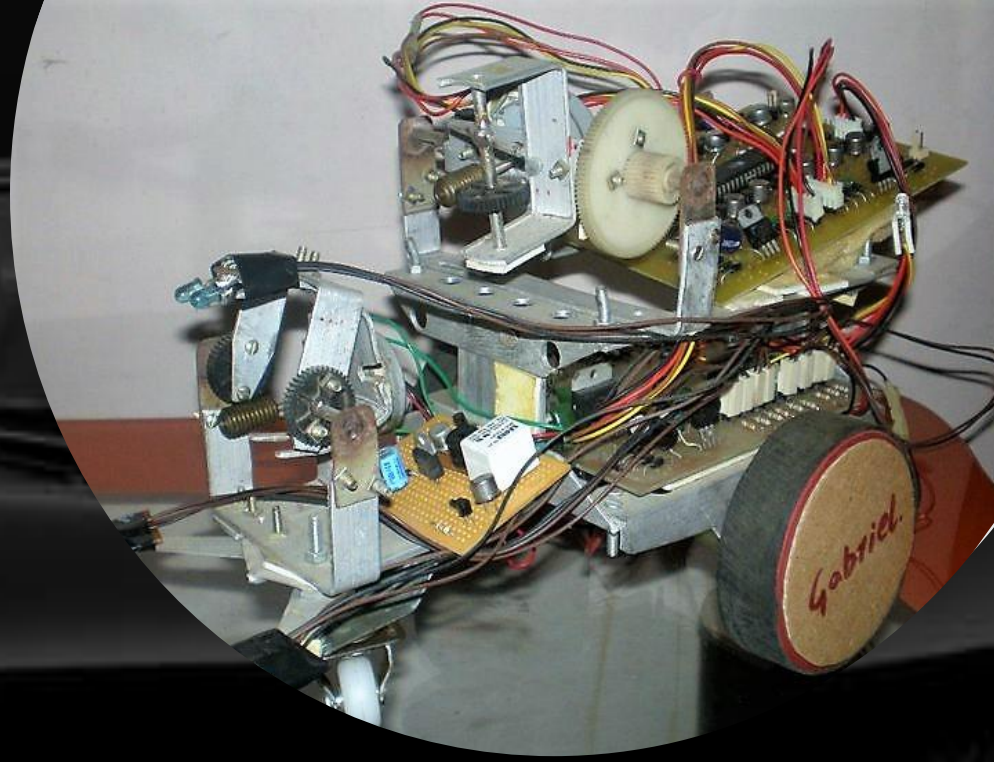
Affiliated By

Agenda

- Registration and drinks 6:00 pm
- Opening - 6:15 pm
- Introduction - Sherin Mathew - 6:20 pm
  - ML Introduction and demo later (10 min max)
- ML with Spark on Databricks- Fahd Rafi- 6:30pm
- AI Demo - Fahd - 7:00 pm (20 max)
- AI Forum - Open discussion 7:20pm
  - Questions and debate - Three questions - 5 mins each
- Break - Food and drinks 7:35pm
- Networking Open forum - Demos and Networking - 8:00pm

About me

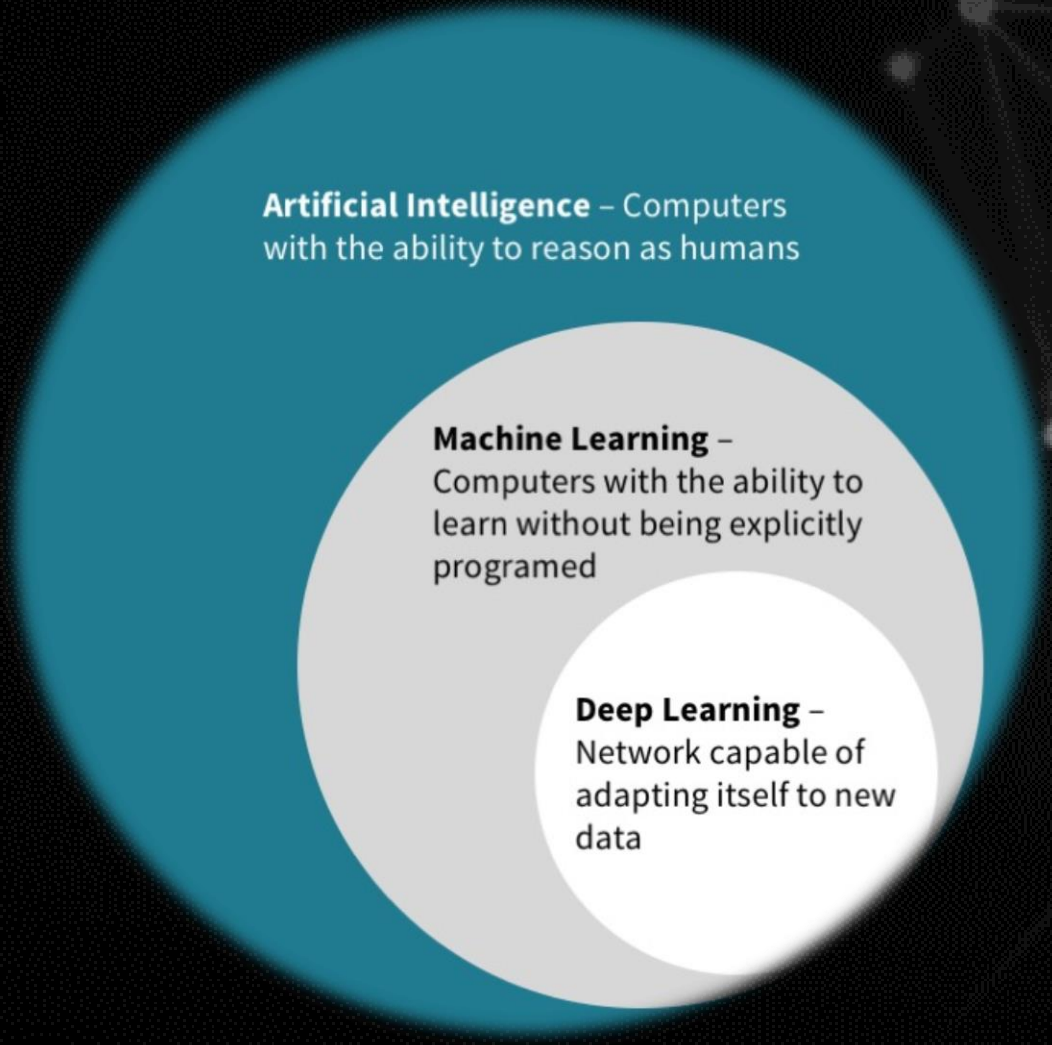
- Principal Consultant @Simpsons , Microsoft Gold Partner
- Azure Architecture - Enterprise BI Solution
- Passionate Data architect and “Azure Addict”
- MCP, MSc and B.Eng. MSP, Agile, Prince2
- Sports: Cricket
- Hobbies: Dancing, Recycled-Art
- Learning: To fly Helicopter – G2 Cabri
- Other interests: Classic Sports car
- Twitter: @smdisrupt



AI User Group

- Our Purpose
- Our Goals
- Announcements
  - Next Event
  - Entry on the website
- Roadmap –Partners, Speakers, Bloggers
- Feedback
- #AIUserGroup @aitechug

AI-Gold-Rush



- 1956, the term “AI” coined by scientist John McCarthy at Dartmouth College
- Around 1980 ML starts to flourish
  - IBM’s Deep Blue computer beats Russian chess master Garry Kasparov in 1997
  - IBM’s Watson beat human players on Jeopardy in 2011 utilizing ML
- 2010 – Deep Learning flourishes
  - 2017 – Densenet is announced
- Today everyone is chasing AI as the new Gold-Rush

Convergence

Data

Cloud, API

Internet

As a Service,
Blockchain, Disruption

Technology

GPU, TPU, Quantum

Automation

Warehouse, Factories,
Transport

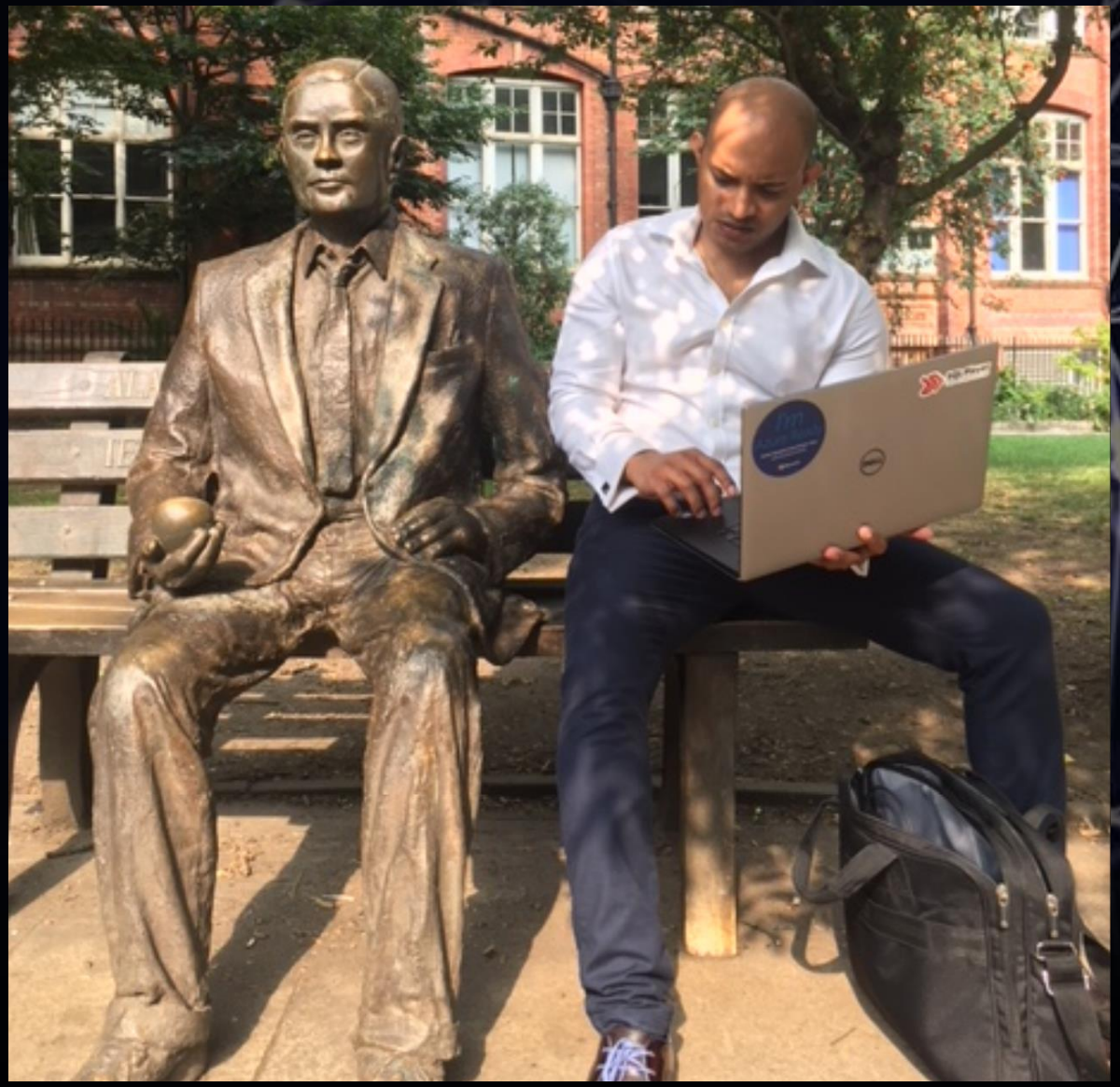
Intelligence –
Driverless, Smart
Gadgets

Communication

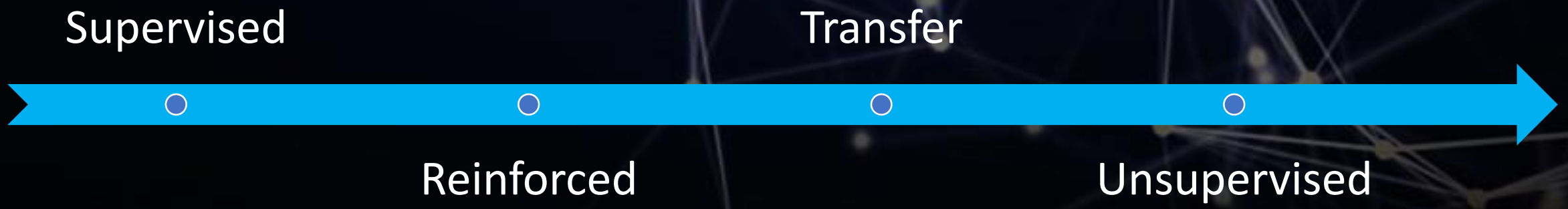
– Devices, Mobile Data

Reality vs Expectation

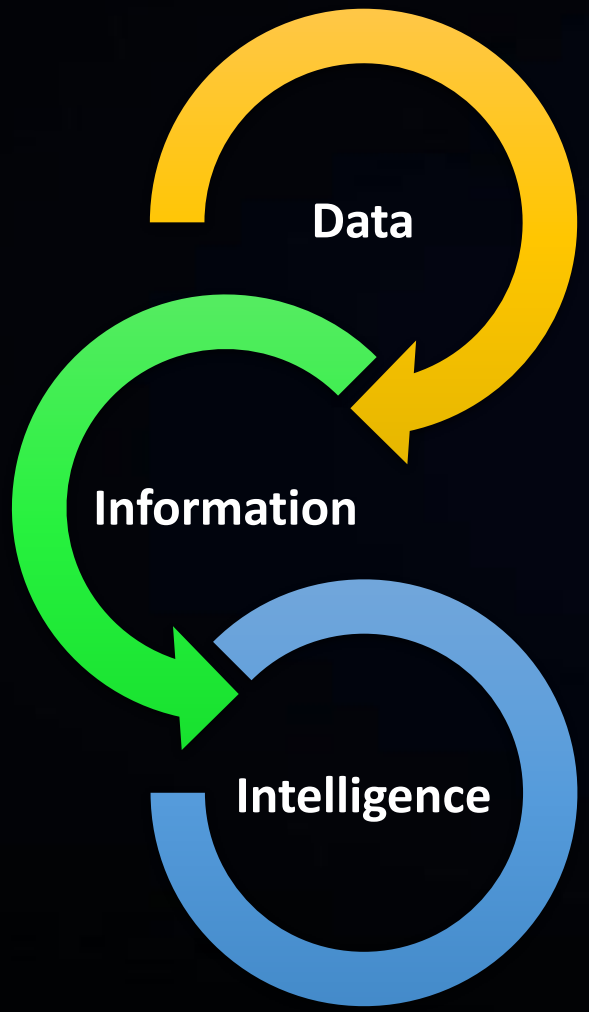
Don’t worry, we are okay for now!

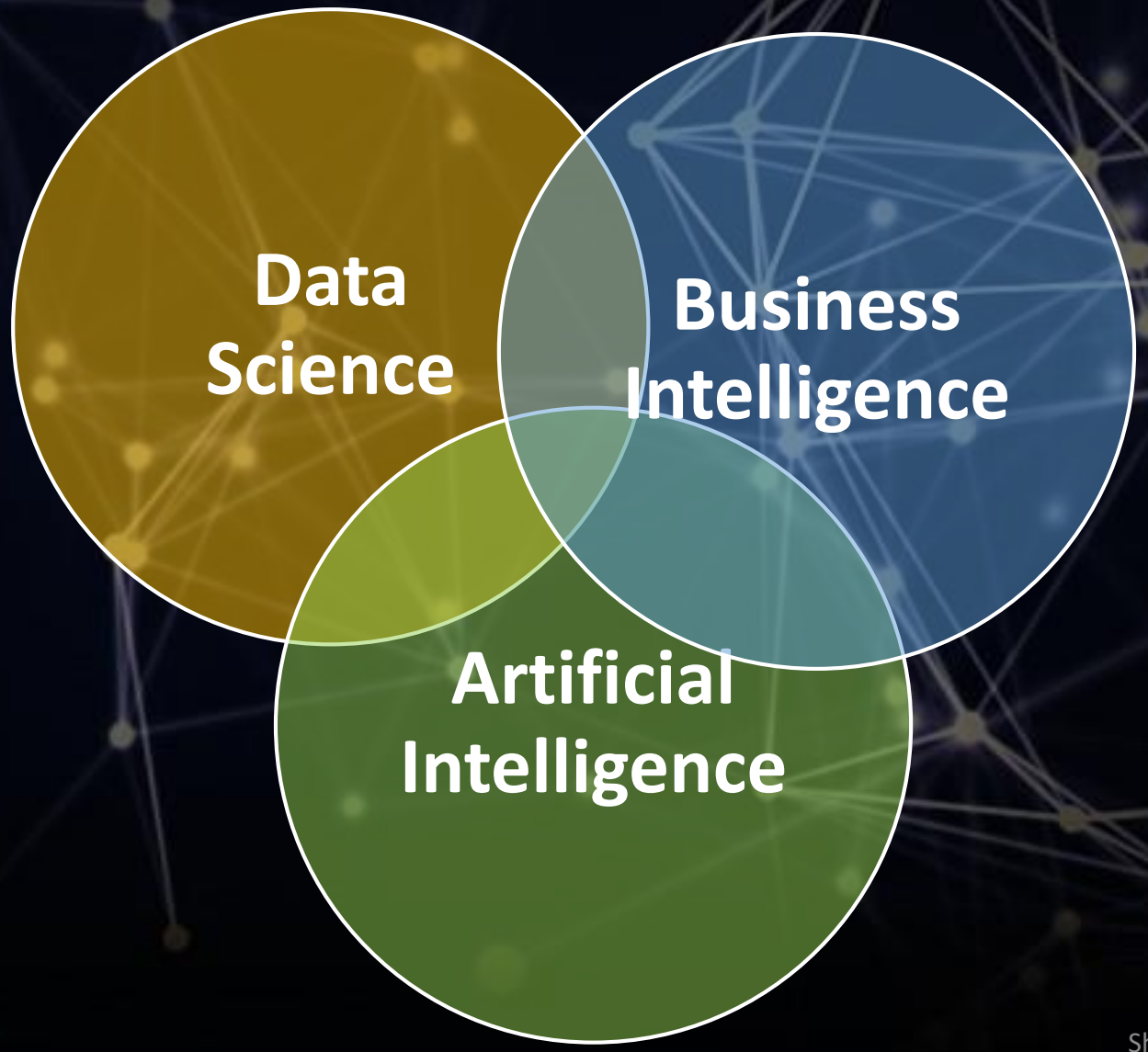


The AI Mechanism



Talk is cheap, show me your intelligence!





Experiment – how do we learn?

Machine Learning in ML Studio

https://studio.azureml.net

Guest Access Workspace: Free trial access without logging in.
Free Workspace: Free persisted access, no Azure subscription needed.
Standard Workspace: Full access with SLA under an Azure subscription.

Anomaly Detection
- One-class Support Vector Machine
- Principal Component Analysis-based Anomaly Detection
- Time Series Anomaly Detection*

Classification

Two-class Classification
- Averaged Perceptron
- Bayes Point Machine
- Boosted Decision Tree
- Decision Forest
- Decision Jungle
- Logistic Regression
- Neural Network
- Support Vector Machine

Multi-class Classification
- Decision Forest
- Decision Jungle
- Logistic Regression
- Neural Network
- One-vs-all

Clustering
- K-means Clustering

Recommendation
- Matchbox Recommender

Regression
- Bayesian Linear Regression
- Boosted Decision Tree
- Decision Forest
- Fast Forest Quantile Regression
- Linear Regression
- Neural Network Regression
- Ordinal Regression
- Poisson Regression

Statistical Functions
- Descriptive Statistics
- Hypothesis Testing T-Test
- Linear Correlation
- Probability Function Evaluation

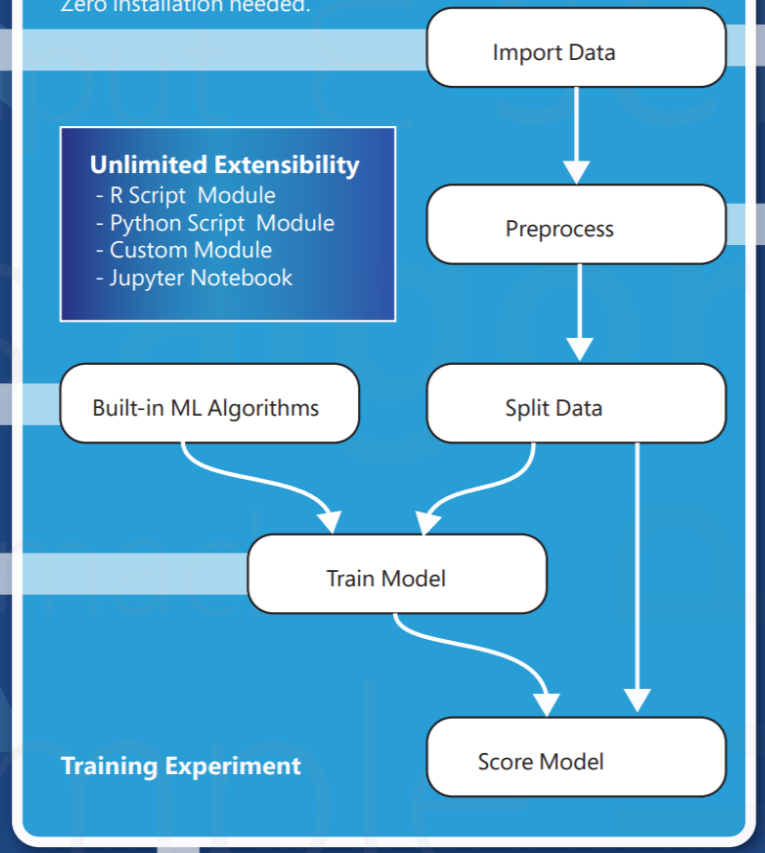
Text Analytics
- Feature Hashing
- Named Entity Recognition
- Vowpal Wabbit

Computer Vision
- OpenCV Library

Data/Model Visualization
- Scatterplots
- Bar Charts
- Box plots
- Histogram
- R and Python Plotting Libraries
- REPL with Jupyter Notebook
- ROC, Precision/Recall, Lift
- Confusion Matrix
- Decision Tree*

Training
- Cross Validation
- Retraining
- Parameter Sweep

Cross browser drag & drop ML workflow designer.
Zero installation needed.



Unlimited Extensibility
- R Script Module
- Python Script Module
- Custom Module
- Jupyter Notebook

Data Source
- Azure Blob Storage
- Azure SQL DB
- Azure SQL DW*
- Azure Table
- Desktop Direct Upload
- Hadoop Hive Query
- Manual Data Entry
- OData Feed
- On-prem SQL Server*
- Web URL (HTTP)

Data Format
- ARFF
- CSV
- SVMLight
- TSV
- Excel
- ZIP

Data Preparation
- Clean Missing Data
- Clip Outliers
- Edit Metadata
- Feature Selection
- Filter
- Learning with Counts
- Normalize Data
- Partition and Sample
- Principal Component Analysis
- Quantize Data
- SQLite Transformation
- Synthetic Minority Oversampling Technique

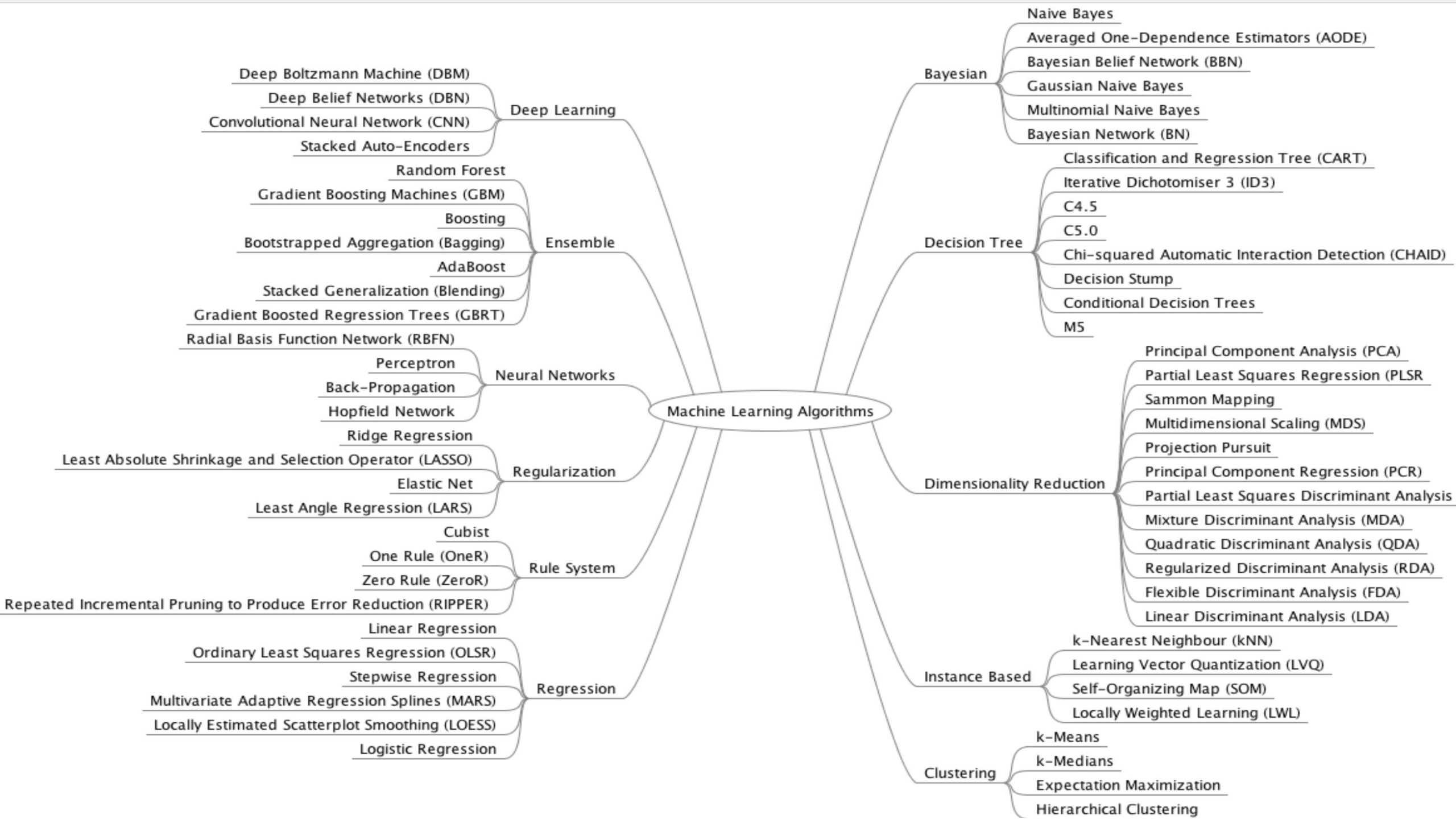
Enterprise Grade Cloud Service
- SLA: 99.95% Guaranteed Up-time
- Azure AD Authentication
- Compute at Large Scale
- Multi-geo Availability
- Regulatory Compliance*

One-click Operationalization

Predictive Experiment

Make Prediction with Elastic APIs
- Request-Response Service (RRS)
- Batch Execution Service (BES)
- Retraining API

Community
- Gallery (http://gallery.azureml.net)
- Samples & Templates
- Workspace Sharing and Collaboration
- Live Chat & MSDN Forum Support


Machine Learning Algorithms
Bayesian
Naive Bayes
Averaged One-Dependence Estimators (AODE)
Bayesian Belief Network (BBN)
Gaussian Naive Bayes
Multinomial Naive Bayes
Bayesian Network (BN)
Decision Tree
Classification and Regression Tree (CART)
Iterative Dichotomiser 3 (ID3)
C4.5
C5.0
Chi-squared Automatic Interaction Detection (CHAID)
Decision Stump
Conditional Decision Trees
M5
Dimensionality Reduction
Principal Component Analysis (PCA)
Partial Least Squares Regression (PLSR
Sammon Mapping
Multidimensional Scaling (MDS)
Projection Pursuit
Principal Component Regression (PCR)
Partial Least Squares Discriminant Analysis
Mixture Discriminant Analysis (MDA)
Quadratic Discriminant Analysis (QDA)
Regularized Discriminant Analysis (RDA)
Flexible Discriminant Analysis (FDA)
Linear Discriminant Analysis (LDA)
Instance Based
k-Nearest Neighbour (kNN)
Learning Vector Quantization (LVQ)
Self-Organizing Map (SOM)
Locally Weighted Learning (LWL)
Clustering
k-Means
k-Medians
Expectation Maximization
Hierarchical Clustering
Deep Learning
Deep Boltzmann Machine (DBM)
Deep Belief Networks (DBN)
Convolutional Neural Network (CNN)
Stacked Auto-Encoders
Ensemble
Random Forest
Gradient Boosting Machines (GBM)
Boosting
Bootstrapped Aggregation (Bagging)
AdaBoost
Stacked Generalization (Blending)
Gradient Boosted Regression Trees (GBRT)
Neural Networks
Radial Basis Function Network (RBFN)
Perceptron
Back-Propagation
Hopfield Network
Regularization
Ridge Regression
Least Absolute Shrinkage and Selection Operator (LASSO)
Elastic Net
Least Angle Regression (LARS)
Rule System
Cubist
One Rule (OneR)
Zero Rule (ZeroR)
Repeated Incremental Pruning to Produce Error Reduction (RIPPER)
Regression
Linear Regression
Ordinary Least Squares Regression (OLSR)
Stepwise Regression
Multivariate Adaptive Regression Splines (MARS)
Locally Estimated Scatterplot Smoothing (LOESS)
Logistic Regression

Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning Studio algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.

START

Finding unusual data points

Discovering structure

Predicting categories

Predicting values

Three or more

Two

ANOMALY DETECTION

- **One-class SVM** — >100 features, aggressive boundary
- **PCA-based anomaly detection** — Fast training

CLUSTERING

- **K-means**

MULTI-CLASS CLASSIFICATION

- Fast training, linear model — **Multiclass logistic regression**
- Accuracy, long training times — **Multiclass neural network**
- Accuracy, fast training — **Multiclass decision forest**
- Accuracy, small memory footprint — **Multiclass decision jungle**
- Depends on the two-class classifier, see notes below — **One-v-all multiclass**

REGRESSION

- **Ordinal regression** — Data in rank ordered categories
- **Poisson regression** — Predicting event counts
- **Fast forest quantile regression** — Predicting a distribution
- **Linear regression** — Fast training, linear model
- **Bayesian linear regression** — Linear model, small data sets
- **Neural network regression** — Accuracy, long training time
- **Decision forest regression** — Accuracy, fast training
- **Boosted decision tree regression** — Accuracy, fast training, large memory footprint

TWO-CLASS CLASSIFICATION

- **Two-class SVM** — >100 features, linear model
- **Two-class averaged perceptron** — Fast training, linear model
- **Two-class logistic regression** — Fast training, linear model
- **Two-class Bayes point machine** — Fast training, linear model

- Accuracy, fast training — **Two-class decision forest**
- Accuracy, fast training, large memory footprint — **Two-class boosted decision tree**
- Accuracy, small memory footprint — **Two-class decision jungle**
- >100 features — **Two-class locally deep SVM**
- Accuracy, long training times — **Two-class neural network**

Training material

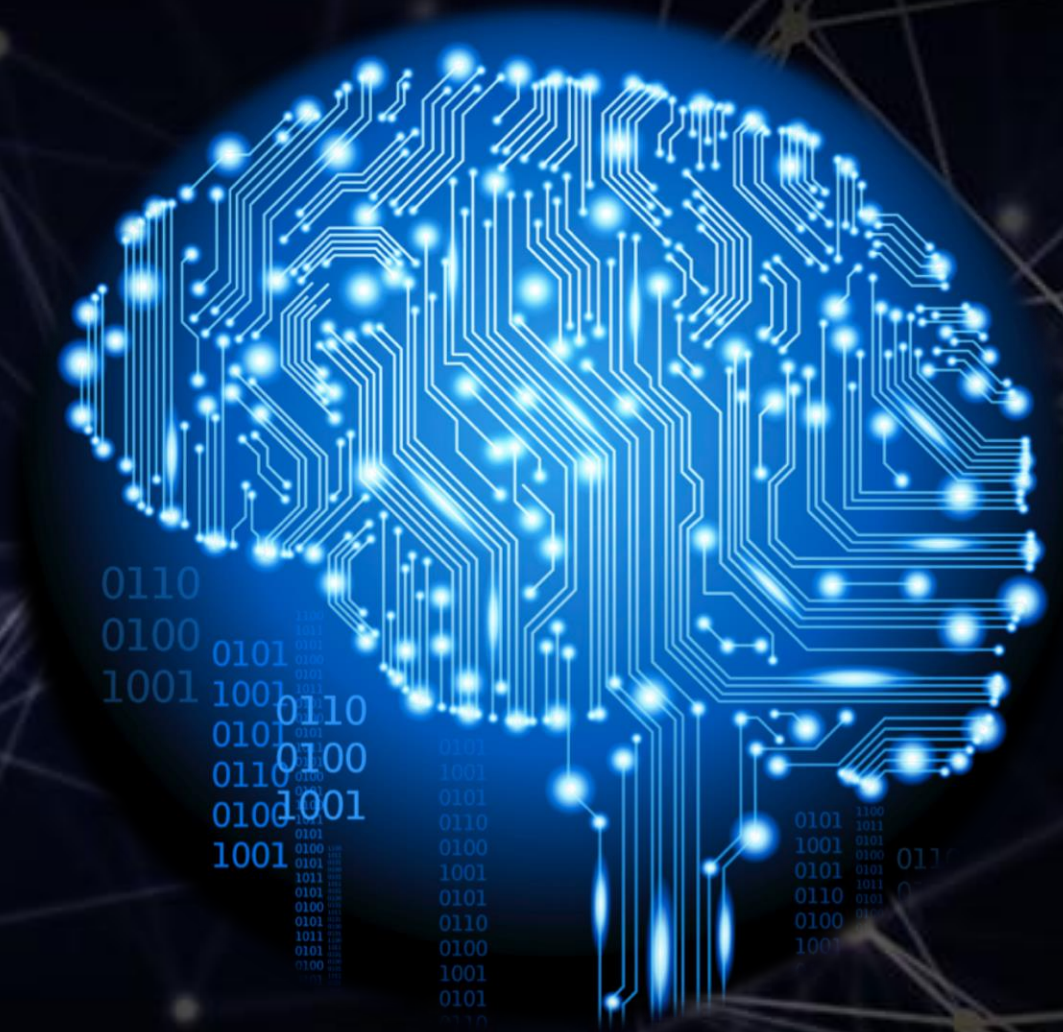
- https://docs.microsoft.com/en-us/azure/machine-learning/service/overview-what-is-azure-ml
- https://docs.microsoft.com/en-us/azure/machine-learning/desktop-workbench/tutorial-classifying-iris-part-1
- https://www.kdnuggets.com/2016/10/clustering-key-terms-explained.html
- https://www.kaggle.com/kanncaa1/machine-learning-tutorial-for-beginners
- https://github.com/rasbt/python-machine-learning-book-2nd-edition
- https://github.com/eriklindernoren/ML-From-Scratch
- https://www.altexsoft.com/blog/datascience/comparing-machine-learning-as-a-service-amazon-microsoft-azure-google-cloud-ai/

Thank you!

Contact:

www.Ai-Tech.uk

AI Tech Uk

@AITechUG #AIUserGroup

Mathewmatics@outlook.com
@smdisrupt