Introduction to Machine Learning

Mohd Aftar bin Abu Bakar

[1]Pusat Pengajian Sains Matematik
Fakulti Sains dan Teknologi
UKM

[2]DELTA
UKM

2017

What is ML?

- A branch of artificial intelligence, concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data.
- The goal is to program computers to use example data or past experience to solve a problem.
- 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)

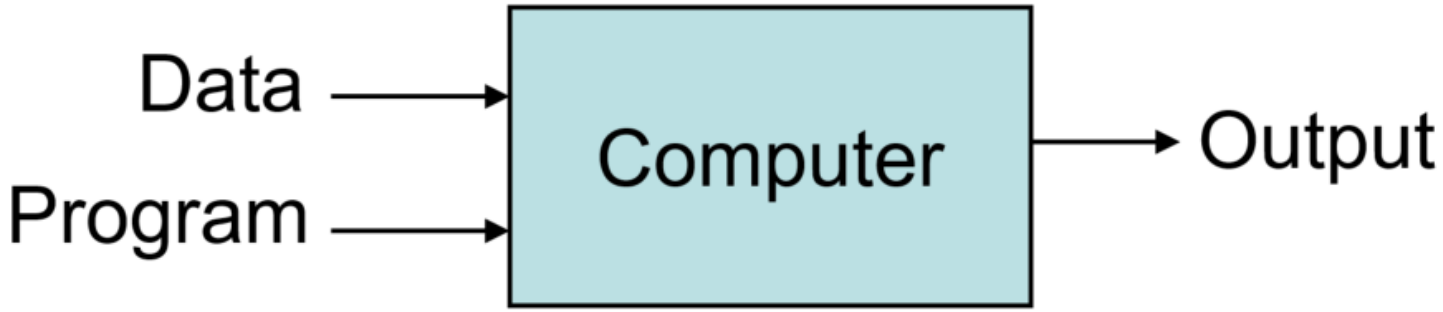
The essence of ML

- A pattern exists
- We cannot pin it down mathematically
- We have data on it

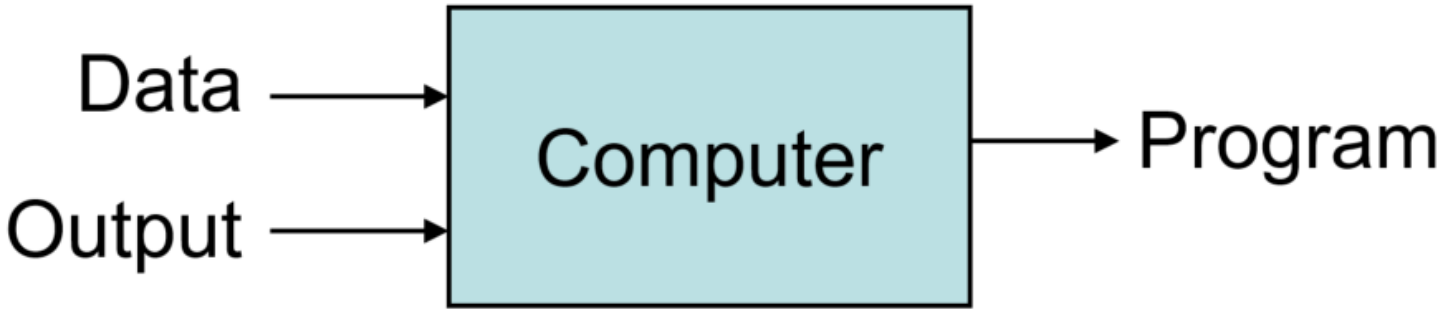
- Automating automation
- Getting computers to program themselves
- Writing software is the bottleneck
- Let the data do the work instead!

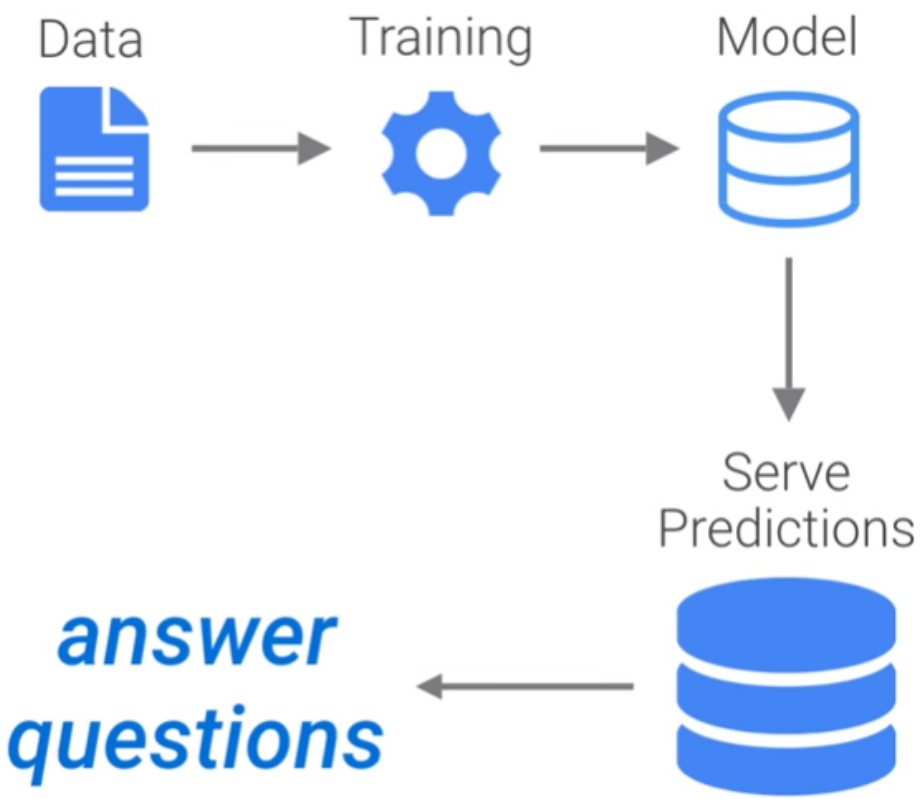
Using data to answer questions

Traditional Programming



Machine Learning



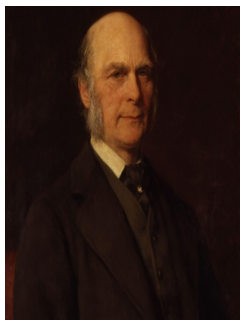
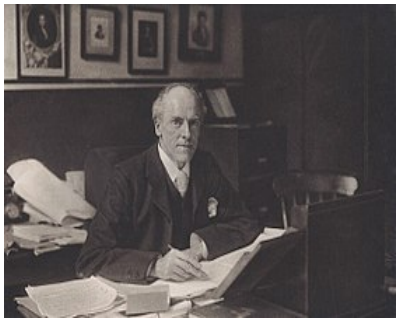
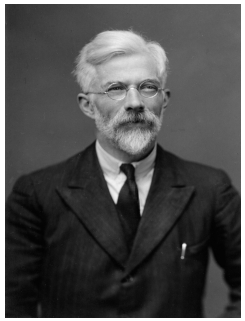

Data
Training
Model
Serve
Predictions
answer
questions

- Applications:
  - analyze past sales data to predict customer behavior,
  - optimize robot behavior so that a task can be completed using minimum resources,
  - extract knowledge from bioinformatics data.

History of ML

- Early 19th Century - Gauss and Legendre independently discovering the method of least squares.
- Sir Francis Galton - Regression to the mean (Tall Parents -¿ Less tall children), correlation. Cousin of Charles Darwin. Coined term Eugenicist. Discouraged low intelligence people from reproducing
- Karl Pearson - student of Galton. Father of mathematical statistics. Eugenicist and racist
- Ronald Fisher - father of modern stats and experimental design. ANOVA. Also Eugenicist and racist



Machine Learning Timeline

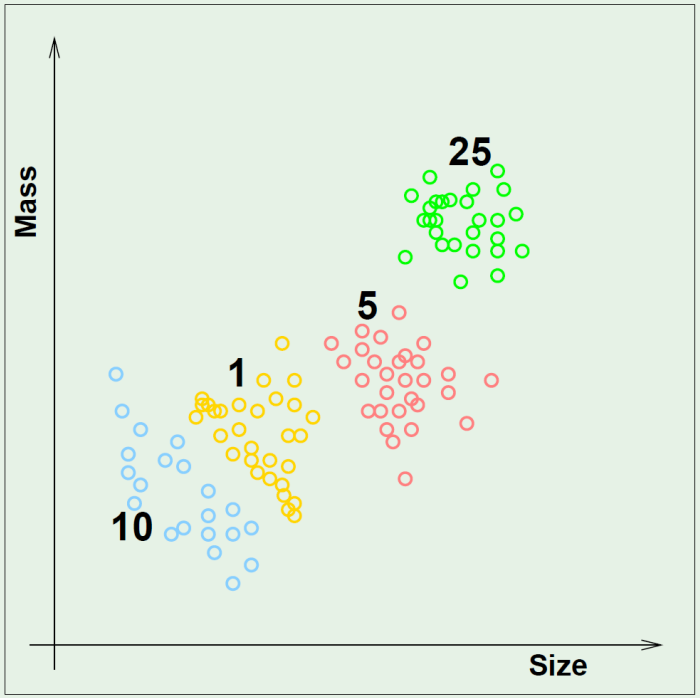
- 1940's: Linear discriminant analysis - First classification method developed by Fisher
- 1950's: Perceptron and Neural Networks - Frank Rosenblatt
- 1960's: Nearest Neighbor, K-means clustering
- 1970's: Logistic regression
- 1980's: Decision Trees and other non-linear methods
- 1990's: Support Vector Machines(Vapnik)
- 2000's: Random Forest (Brieman), Deep Learning (Hinton)

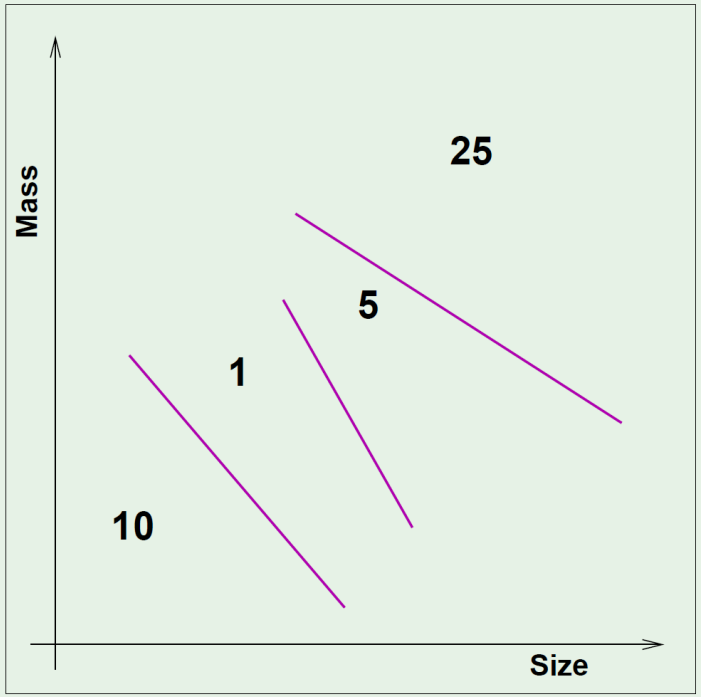
Types of Learning

- **Supervised Learning**
  Training data includes desired outputs.
  Fit a model.
- **Unsupervised Learning**
  Training data does not include desired outputs.
  Clustering.
- **Semi-supervised Learning**
  Training data includes a few desired outputs.
- **Reinforcement Learning**
  Rewards/Punished from sequence of actions. Learns from this and repeats.

Supervised Learning

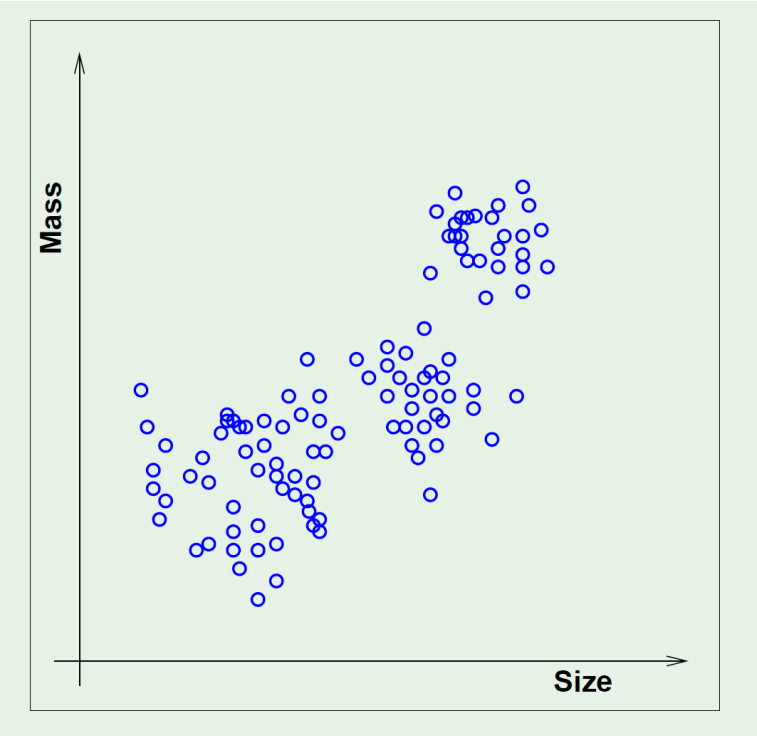
Example: Vending machine - coin recognition





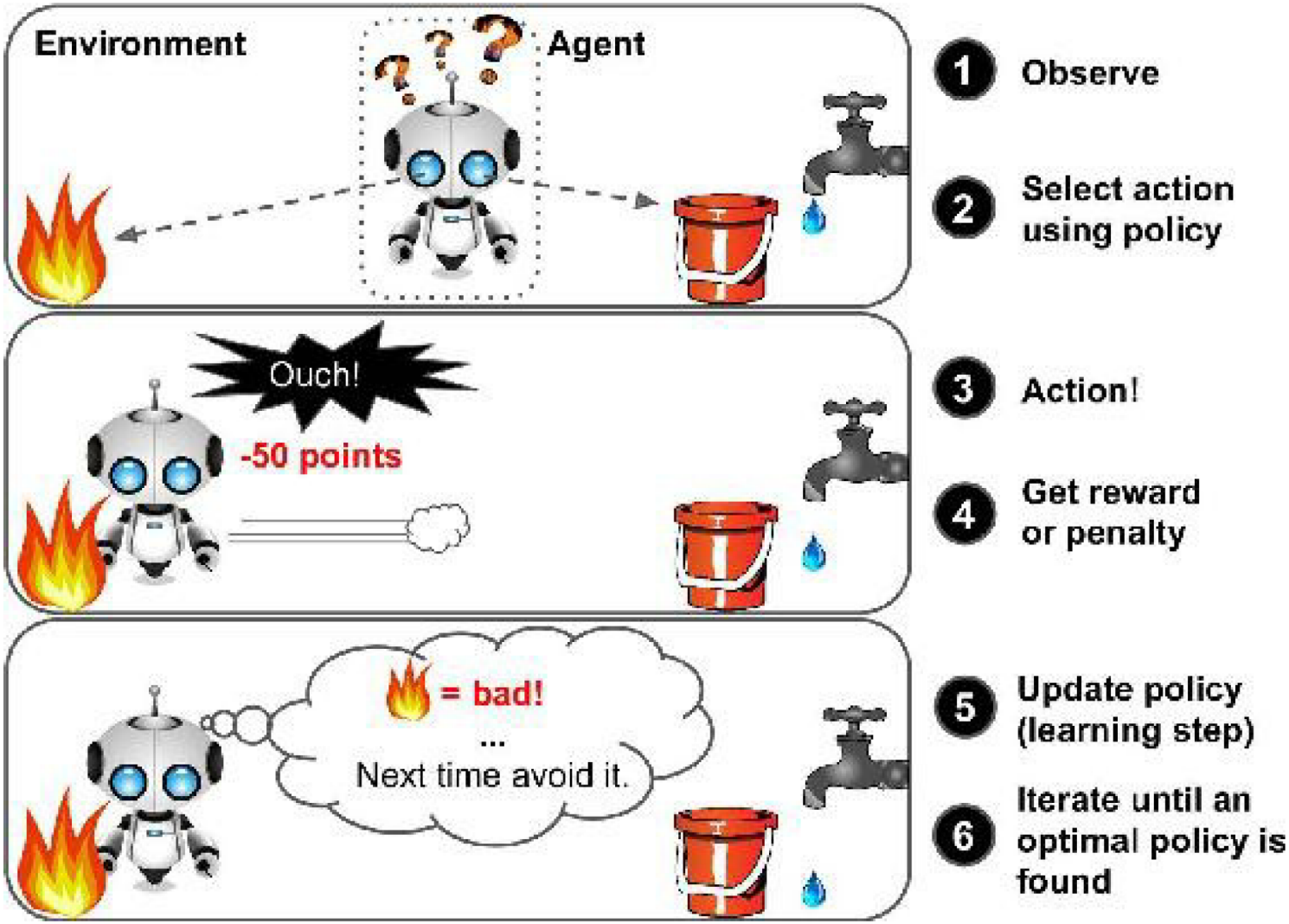
Unsupervised Learning

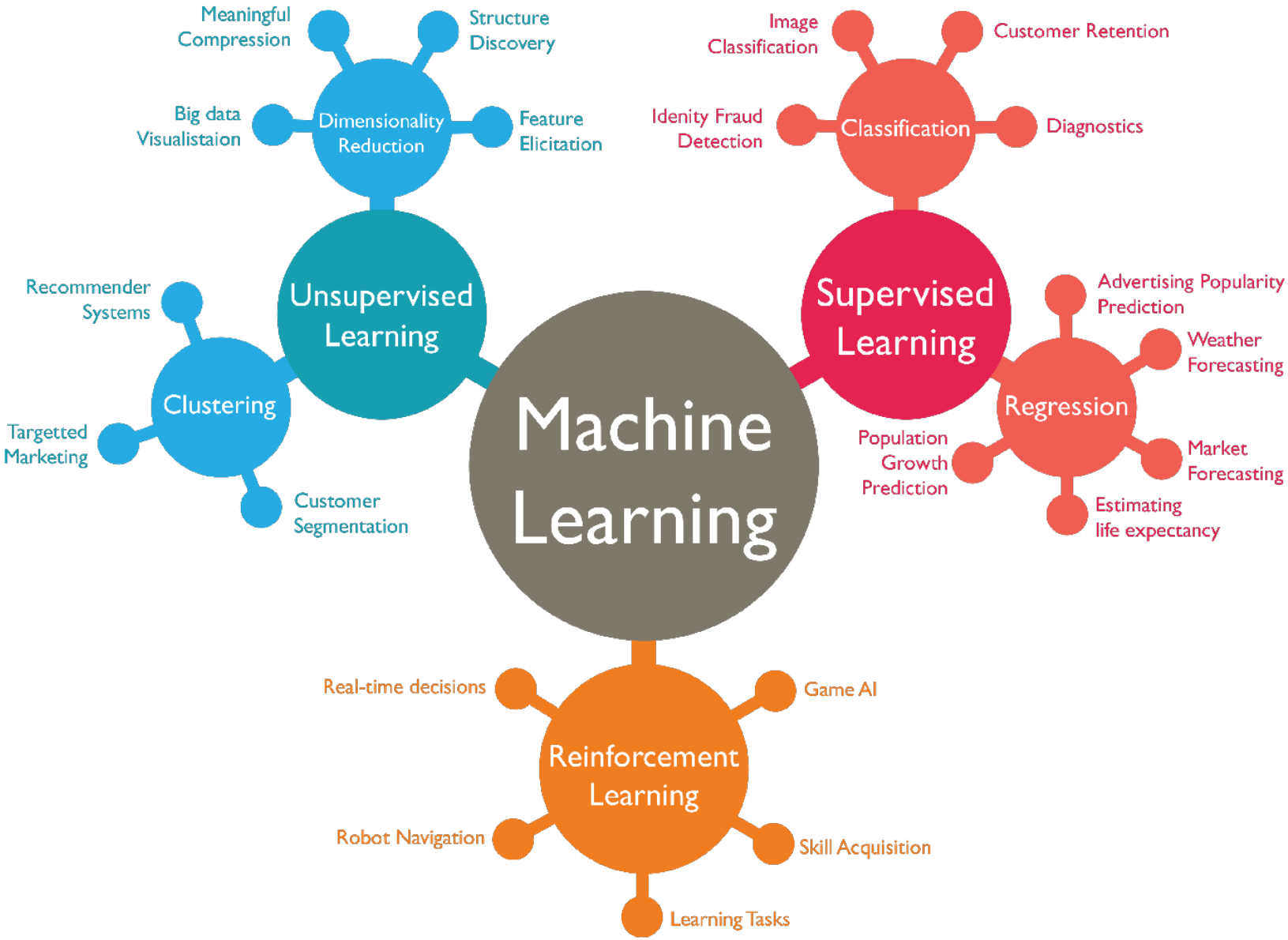
Example: Vending machine - coin recognition



Reinforcement Learning

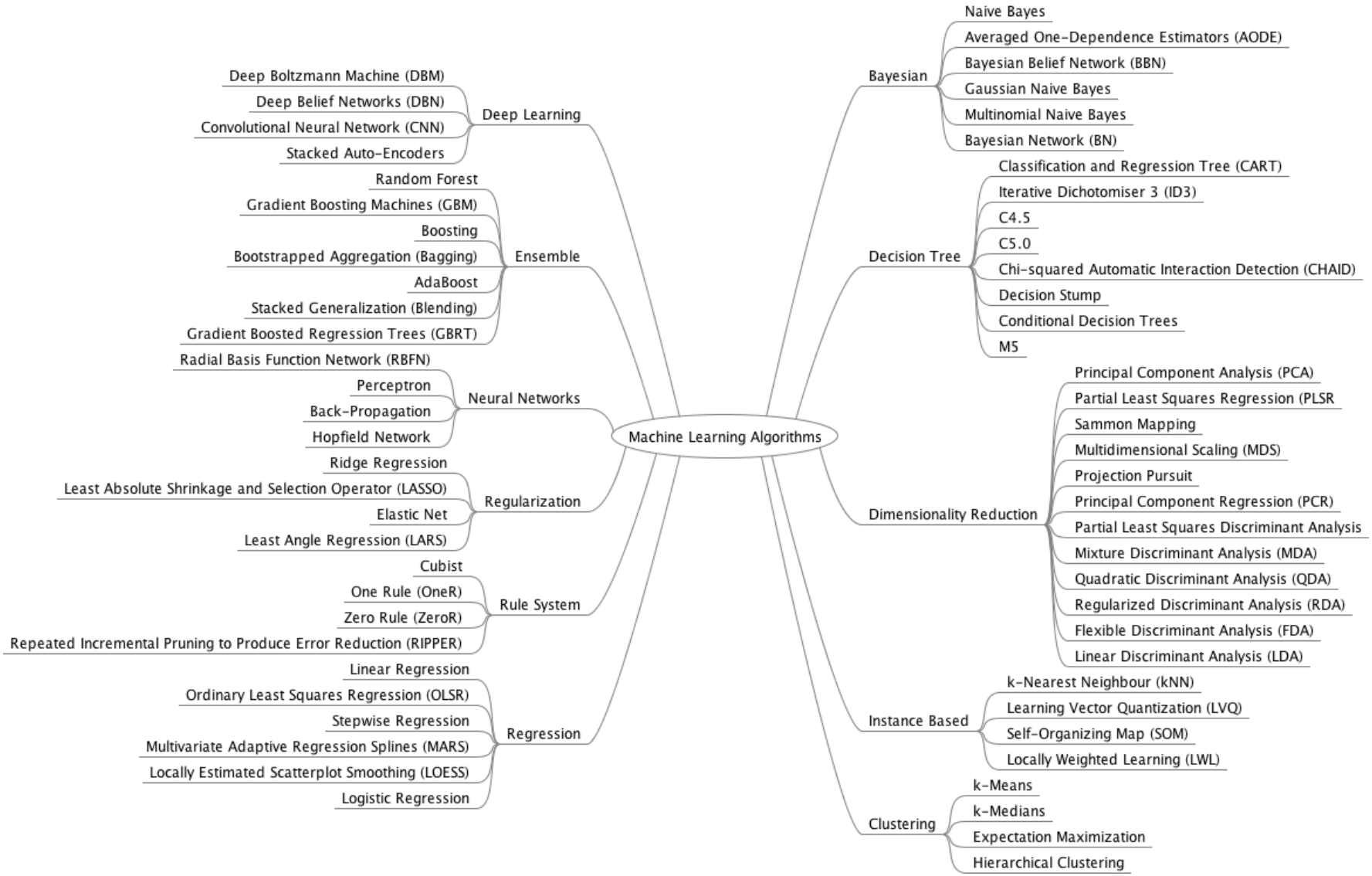
Example:




Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Market Forecasting
Estimating life expectancy
Population Growth Prediction
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

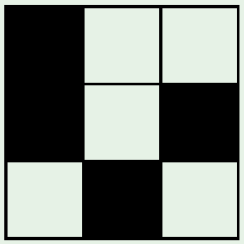
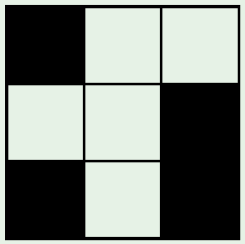
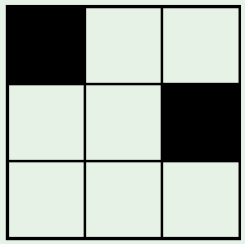
7 Steps of Machine Learning

1. Gather Data
2. Prepare Data
3. Choose Model
4. Training
5. Evaluation
6. Hyperparameter Tuning
7. Prediction

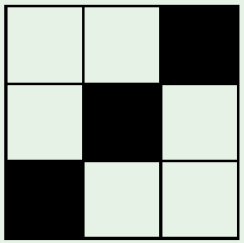
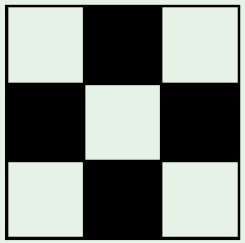
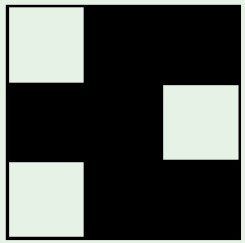

Machine Learning Algorithms
Bayesian
Naive Bayes
Averaged One-Dependence Estimators (AODE)
Bayesian Belief Network (BBN)
Gaussian Naive Bayes
Multinomial Naive Bayes
Bayesian Network (BN)
Decision Tree
Classification and Regression Tree (CART)
Iterative Dichotomiser 3 (ID3)
C4.5
C5.0
Chi-squared Automatic 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

Visual Intro to Machine Learning

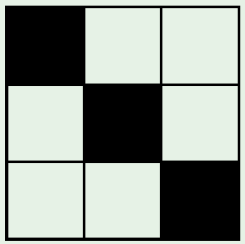
A Learning puzzle

$f = -1$

$f = +1$



$f = ?$