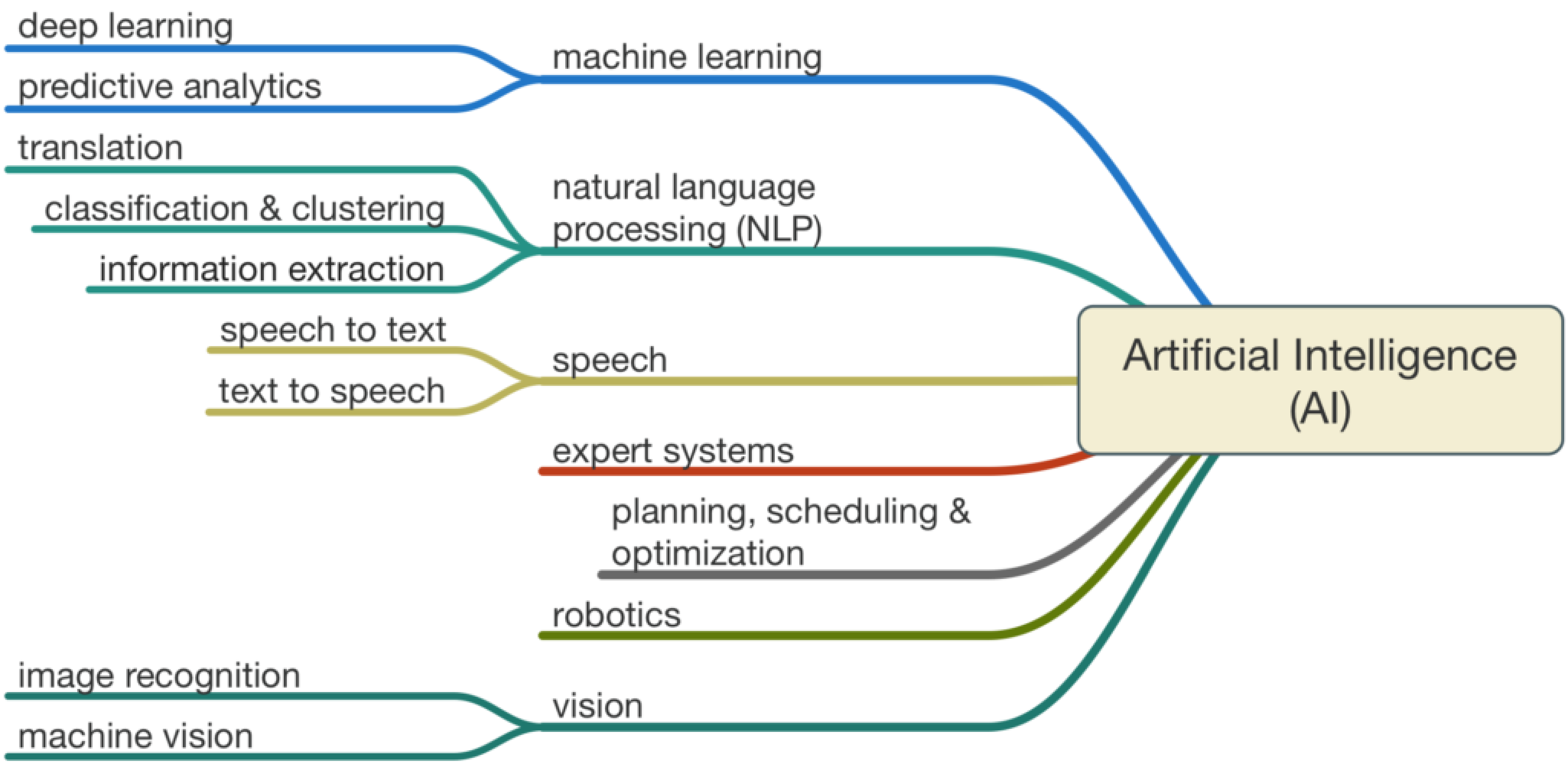
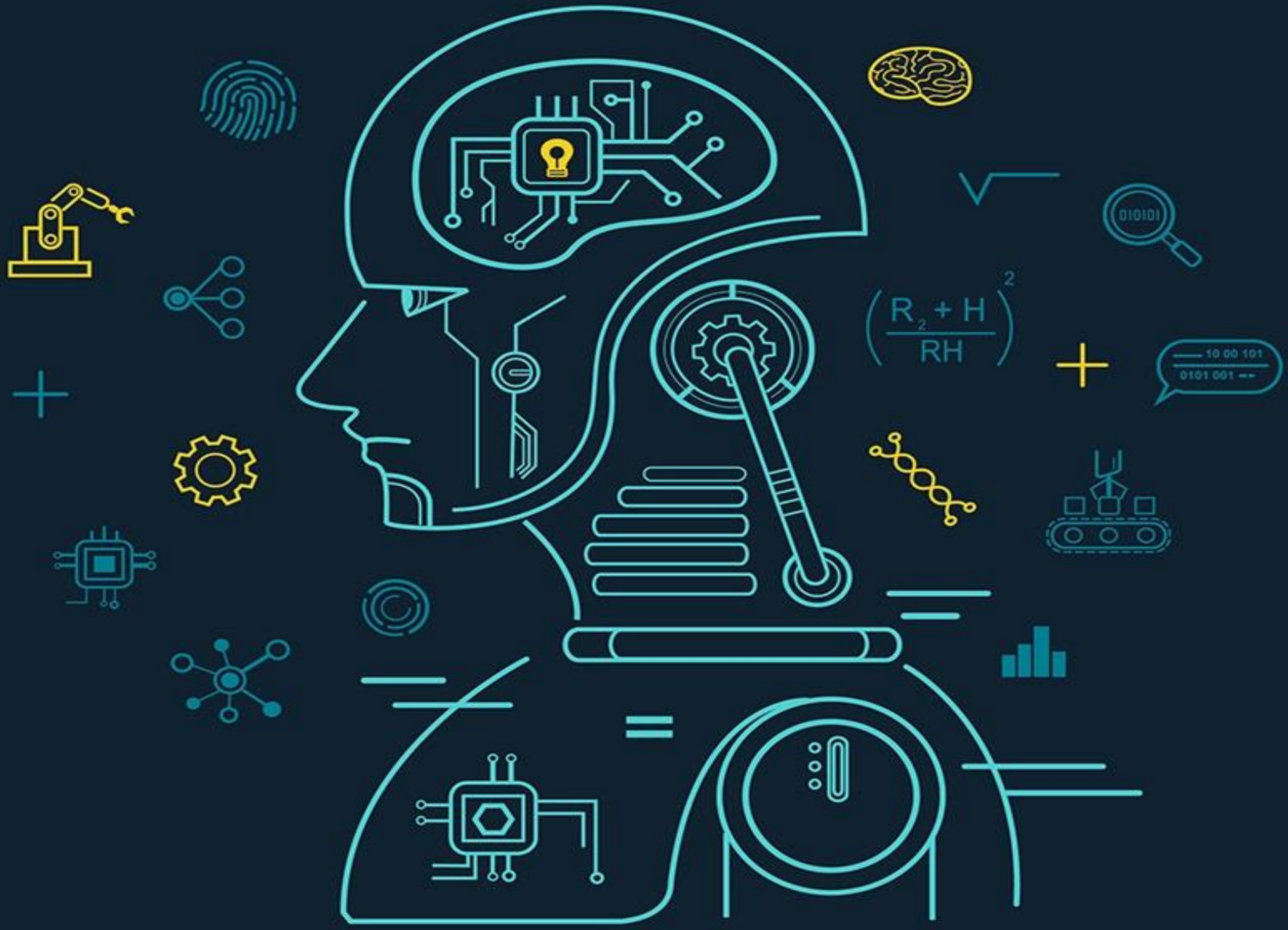
From AI to ML

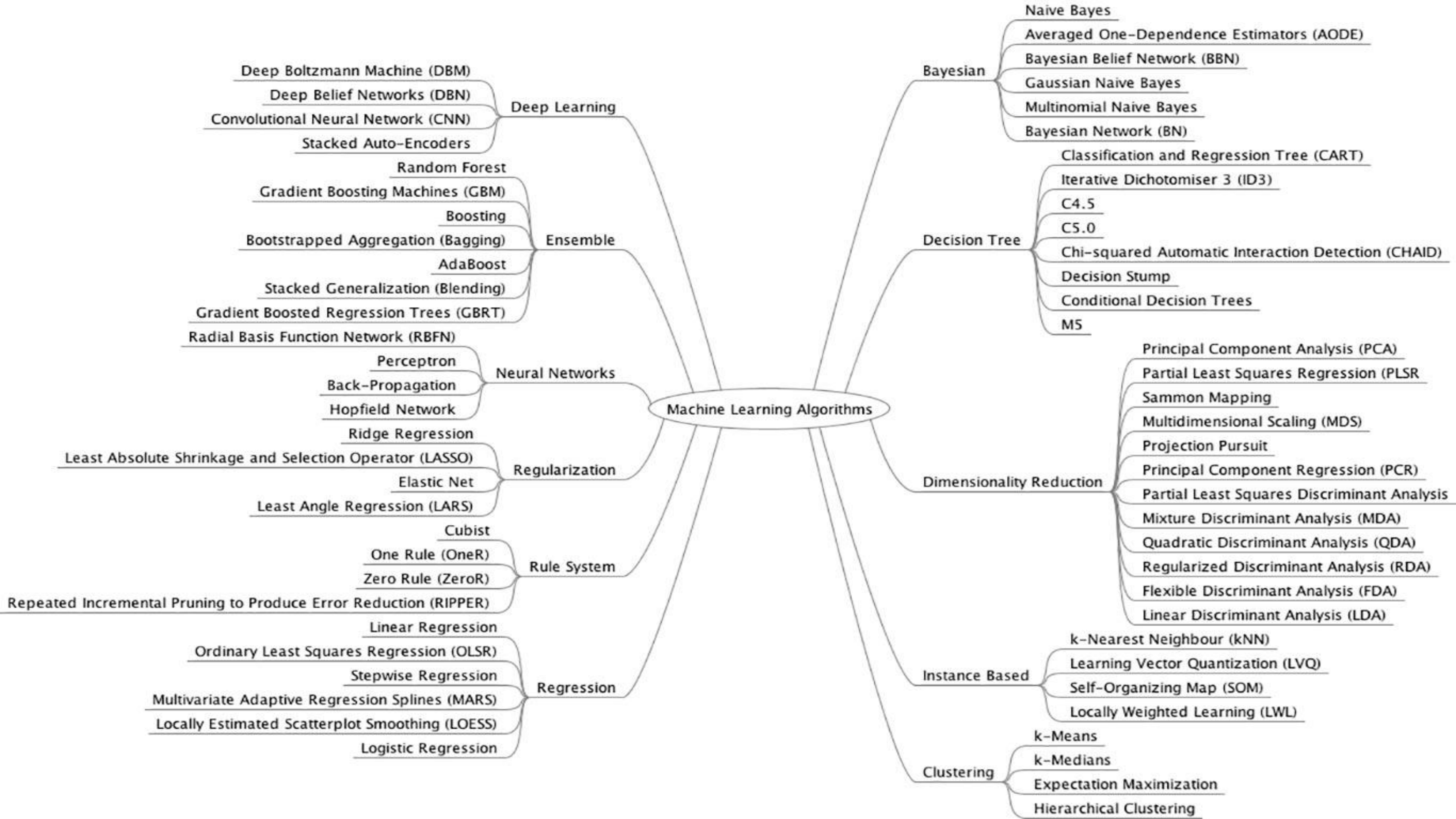
BY MXK

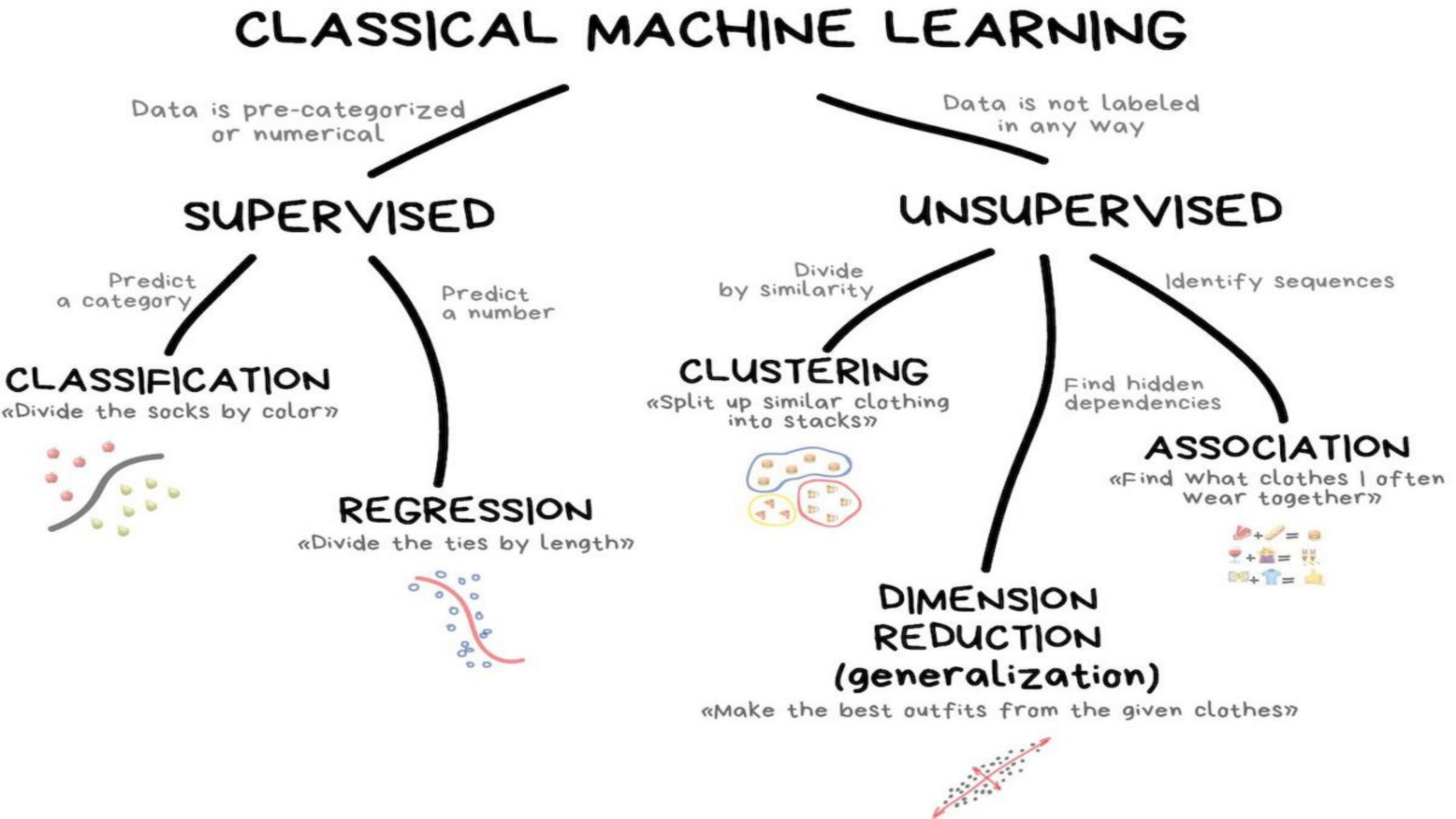

AI
ARTIFICIAL INTELLIGENCE


deep learning
predictive analytics
machine learning
translation
classification & clustering
information extraction
natural language processing (NLP)
speech to text
text to speech
speech
Artificial Intelligence (AI)
expert systems
planning, scheduling & optimization
robotics
image recognition
machine vision
vision



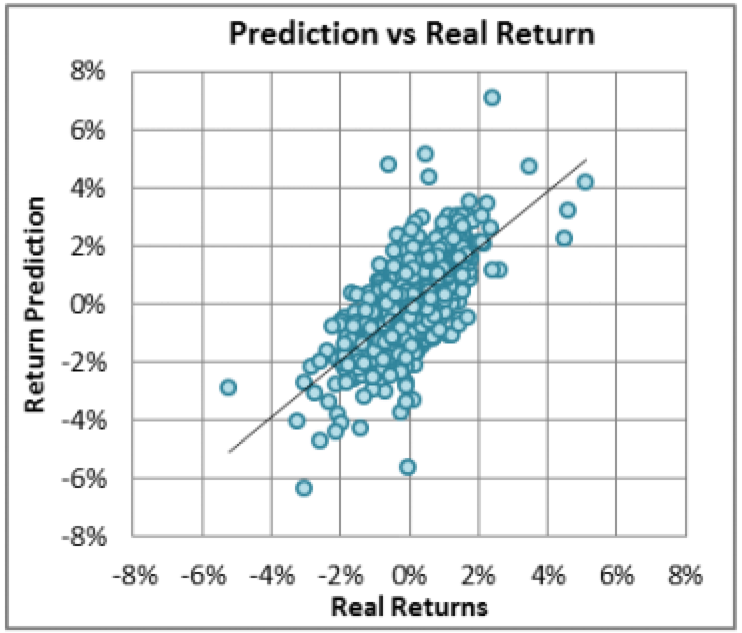
MACHINE LEARNING

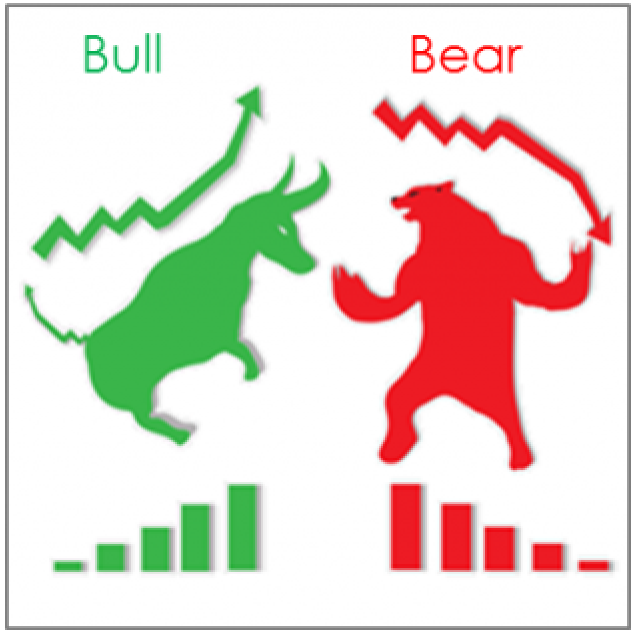

Machine Learning Algorithms
Deep Learning
Deep Boltzmann Machine (DBM)
Deep Belief Networks (DBN)
Convolutional Neural Network (CNN)
Stacked Auto-Encoders
Ensemble
Random Forest
Gradient Boosting Machines (GBM)
Boosting
Bootstrapped Aggregation (Bagging)
AdaBoost
Stacked Generalization (Blending)
Gradient 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
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


CLASSICAL MACHINE LEARNING
Data is pre-categorized or numerical
Data is not labeled in any way
SUPERVISED
UNSUPERVISED
Predict a category
Predict a number
Divide by similarity
Identify sequences
CLASSIFICATION
«Divide the socks by color»
REGRESSION
«Divide the ties by length»
CLUSTERING
«Split up similar clothing into stacks»
Find hidden dependencies
ASSOCIATION
«Find what clothes I often wear together»
DIMENSION REDUCTION (generalization)
«Make the best outfits from the given clothes»

Supervised Machine Learning

Regression vs Classification





4 types of classification:

- Binary
- Multiclass
- Multi-label
- Imbalanced

Binary Classification

Binary classification refers to those classification tasks that have two class labels.

Examples include:
- Email spam detection (spam or not).
- Churn prediction (churn or not).
- Conversion prediction (buy or not).

Popular algorithms that can be used for binary classification include:
- Logistic Regression
- k-Nearest Neighbors
- Decision Trees
- Support Vector Machine
- Naive Bayes

Code for a use case

Multi-Class Classification

Multi-class classification refers to those classification tasks that have more than two class labels.

Examples include:

- Plant species classification.
- Optical character recognition.

Algorithms that are designed for binary classification can be adapted for use for multi-class problems using the following techniques

- One-vs-Rest: Fit one binary classification model for each class vs. all other classes.
- One-vs-One: Fit one binary classification model for each pair of classes.

Popular algorithms that can be used for multi-class classification include:

- Logistic Regression (binary and multi-class).
- Random Forest.

Code for a use case:

Multi-Label Classification

Multi-label classification refers to those classification tasks that have two or more class labels, where one or more class labels may be predicted for each example.

Examples include:
Photo classifications where a given photo may have multiple objects in the scene and a model may predict the presence of multiple known objects in the photo, such as “bicycle,” “apple,” “person,” etc.

Specialized versions of standard classification algorithms can be used:
- Multi-label Decision Trees.
- Multi-label Random Forests.
- Multi-label Gradient Boosting.

Code for a use case:

Imbalanced Classification

Imbalanced classification refers to classification tasks where the number of examples in each class is unequally distributed.

Examples include:

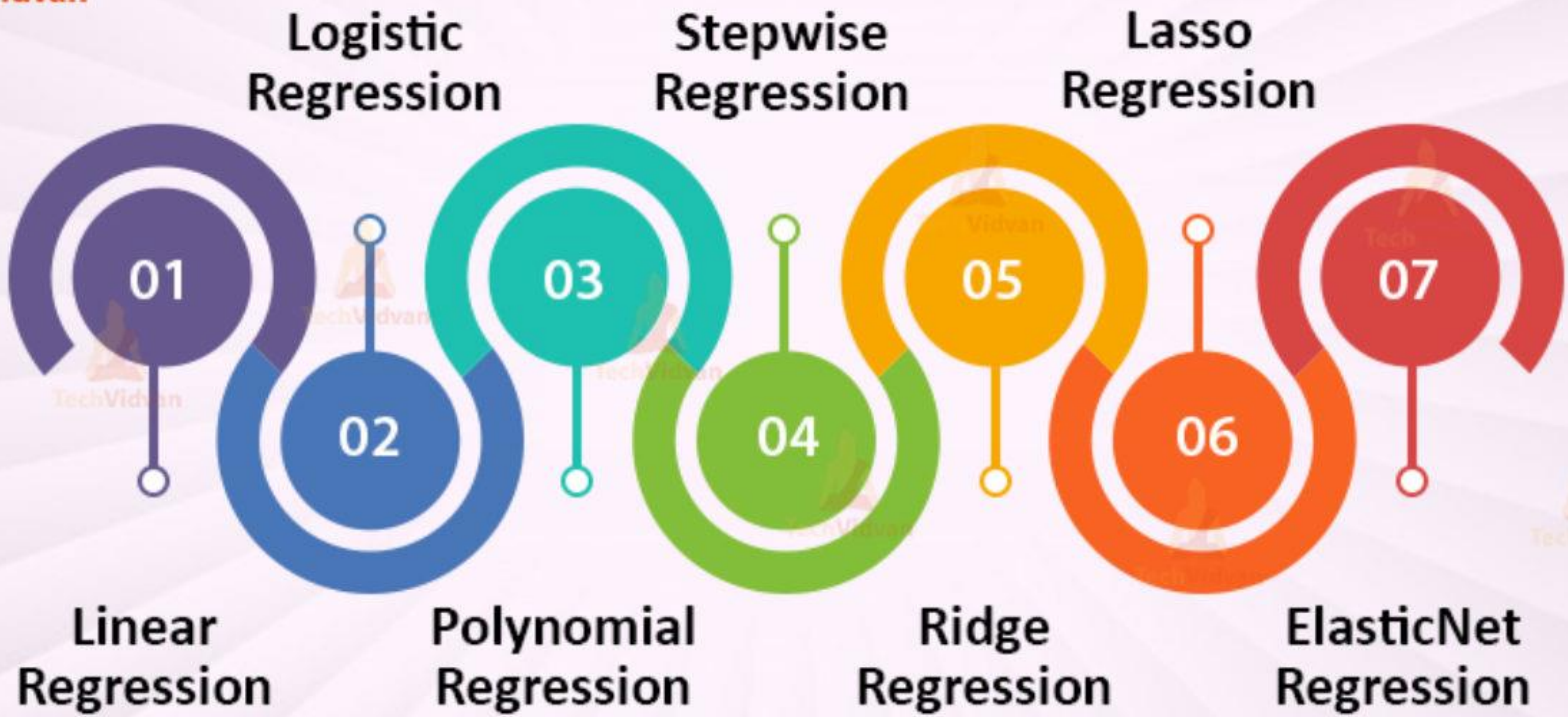
- Fraud detection.
- Outlier detection.
- Medical diagnostic tests.

Specialized modeling algorithms may be used that pay more attention to the minority class when fitting the model on the training dataset:

- Cost-sensitive Logistic Regression.
- Cost-sensitive Decision Trees.
- Cost-sensitive Support Vector Machines.

Code for a use case:

Types of Regression



Types of Regression Part 1

Linear regression
is used for predictive analysis. Linear regression is a linear approach for modeling the relationship between the criterion or the scalar response and the multiple predictors or explanatory variables.
$Y = bX + A$.

Logistic regression
is used when the dependent variable is dichotomous. Logistic regression is used to deal with data that has two possible criterions and the relationship between the criterions and the predictors.
$Y = b0 + b1X1 + b2X2$.

Polynomial regression
is used for curvilinear data. Polynomial regression is fit with the method of least squares. The goal of regression analysis to model the expected value of a dependent variable y in regards to the independent variable x.
$Y = b0 + b0X1 + e$

Types of Regression Part 2

Stepwise regression
is used for fitting regression models with predictive models. It is carried out automatically. With each step, the variable is added or subtracted from the set of explanatory variables. The approaches for stepwise regression are forward selection, backward elimination, and bidirectional elimination.

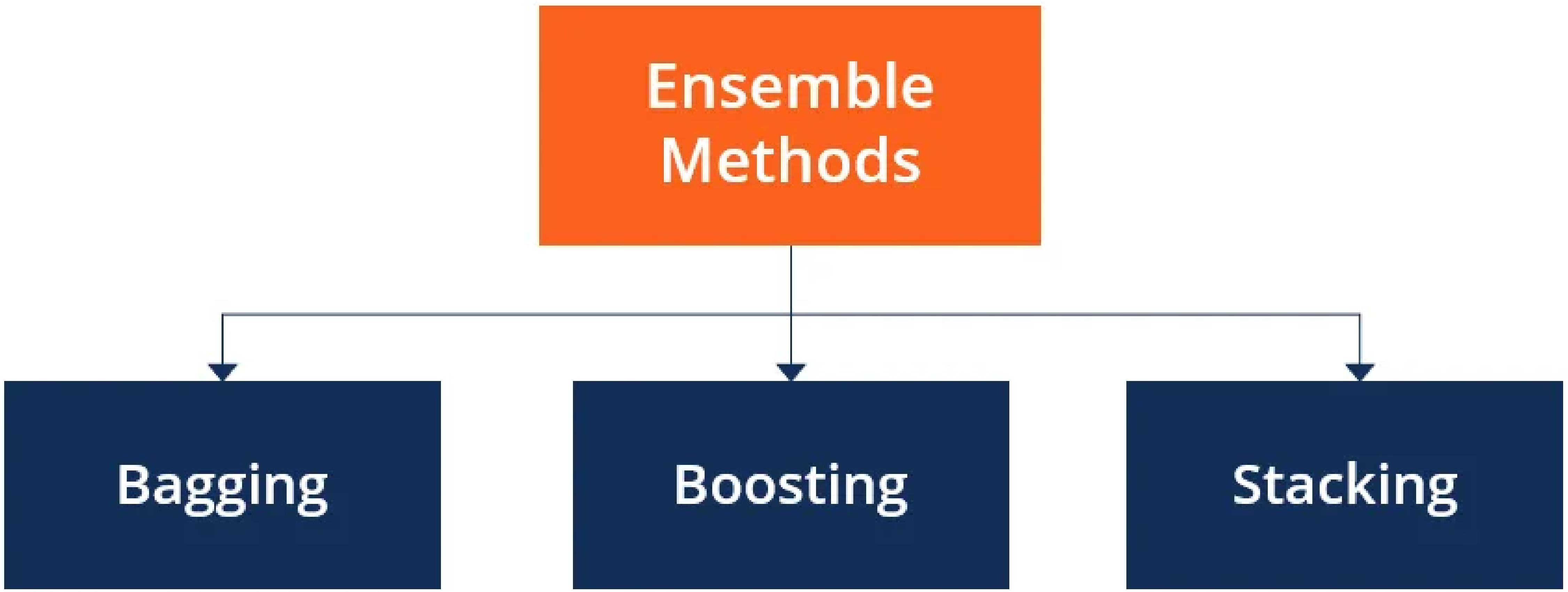
Ridge regression
is a technique for analyzing multiple regression data. When multicollinearity occurs, least squares estimates are unbiased. A degree of bias is added to the regression estimates, and a result, ridge regression reduces the standard errors.

Types of Regression Part 3

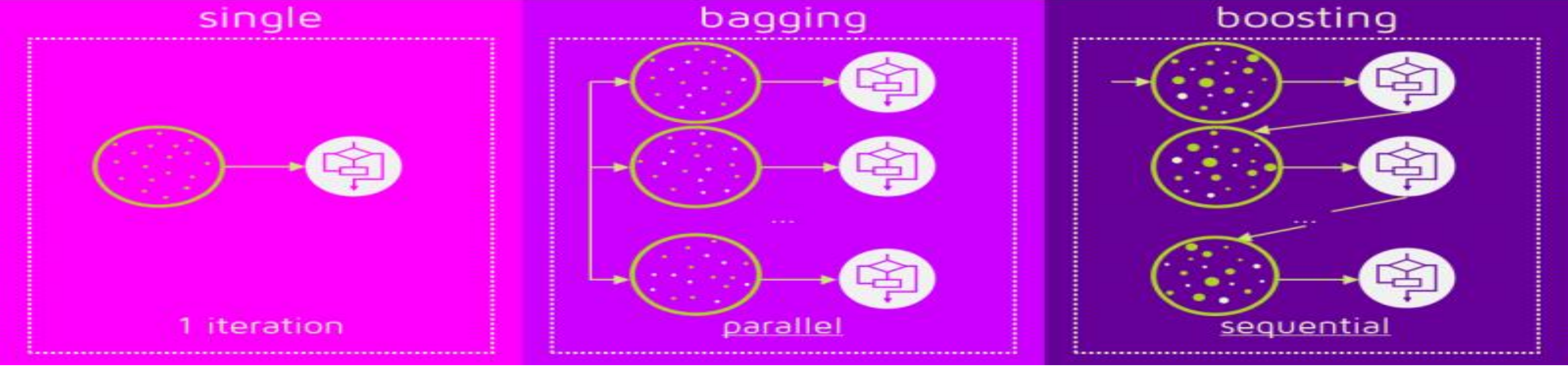
Lasso Regression
is a regression analysis method that performs both variable selection and regularization. Lasso regression uses soft thresholding. Lasso regression selects only a subset of the provided covariates for use in the final model.

ElasticNet Regression
is a regularized regression method that linearly combines the penalties of the lasso and ridge methods. ElasticNet regression is used for support vector machines, metric learning, and portfolio optimization.

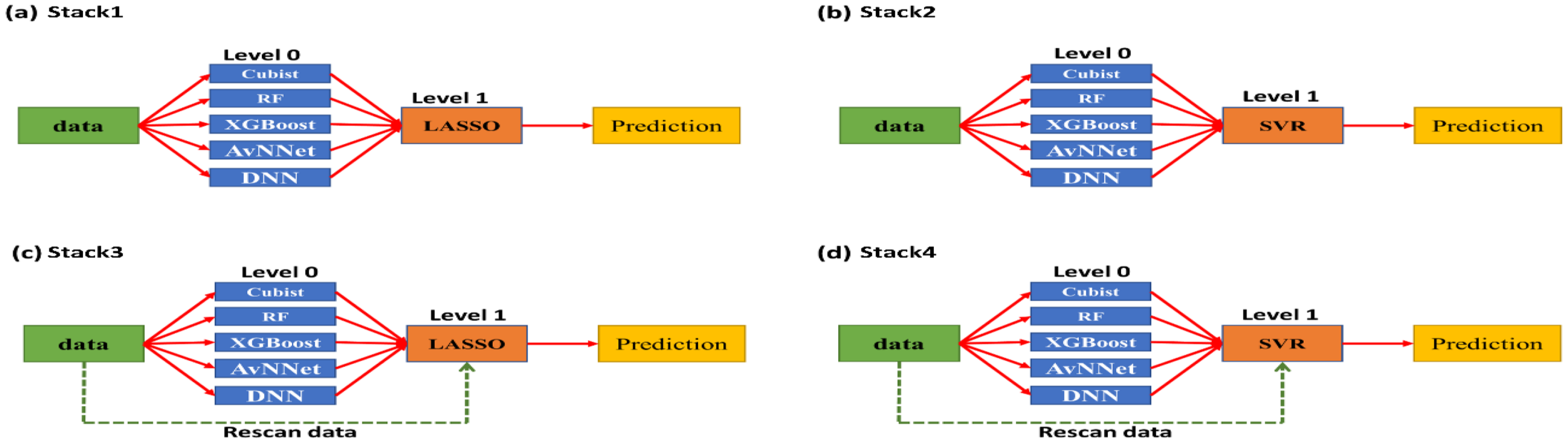
ENSEMBLE LEARNING


Ensemble Methods
Bagging
Boosting
Stacking

	Bagging	Boosting	Stacking
Partitioning of the data into subsets	Random	Giving mis-classified samples higher preference	Various
Goal to achieve	Minimize variance	Increase predictive force	Both
Methods where this is used	Random subspace	Gradient descent	Blending
Function to combine single models	(Weighted) average	Weighted majority vote	Logistic regression



single
1 iteration
bagging
parallel
boosting
sequential



(a) Stack1
Level 0
Cubist
RF
XGBoost
AvNNet
DNN
data
Level 1
LASSO
Prediction
(b) Stack2
Level 0
Cubist
RF
XGBoost
AvNNet
DNN
data
Level 1
SVR
Prediction
(c) Stack3
Level 0
Cubist
RF
XGBoost
AvNNet
DNN
data
Level 1
LASSO
Prediction
Rescan data
(d) Stack4
Level 0
Cubist
RF
XGBoost
AvNNet
DNN
data
Level 1
SVR
Prediction
Rescan data

Colab Time ☺

Kahoot Time ☺