Introduction to Machine Learning

What is machine learning?

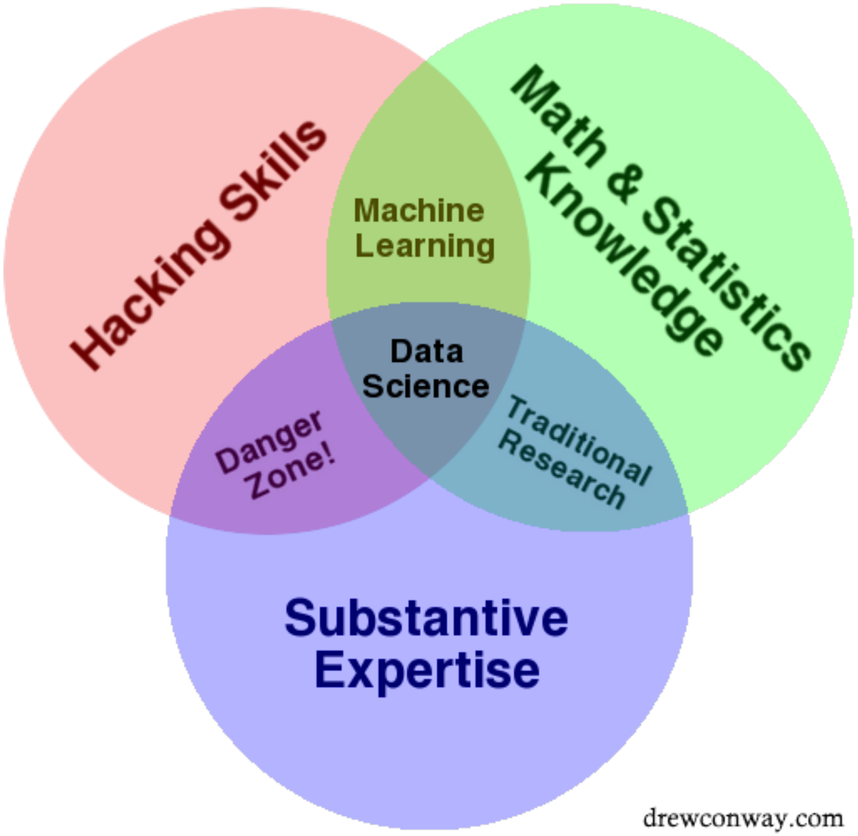
"[A] field of study that gives computers the ability to learn without being explicitly programmed."

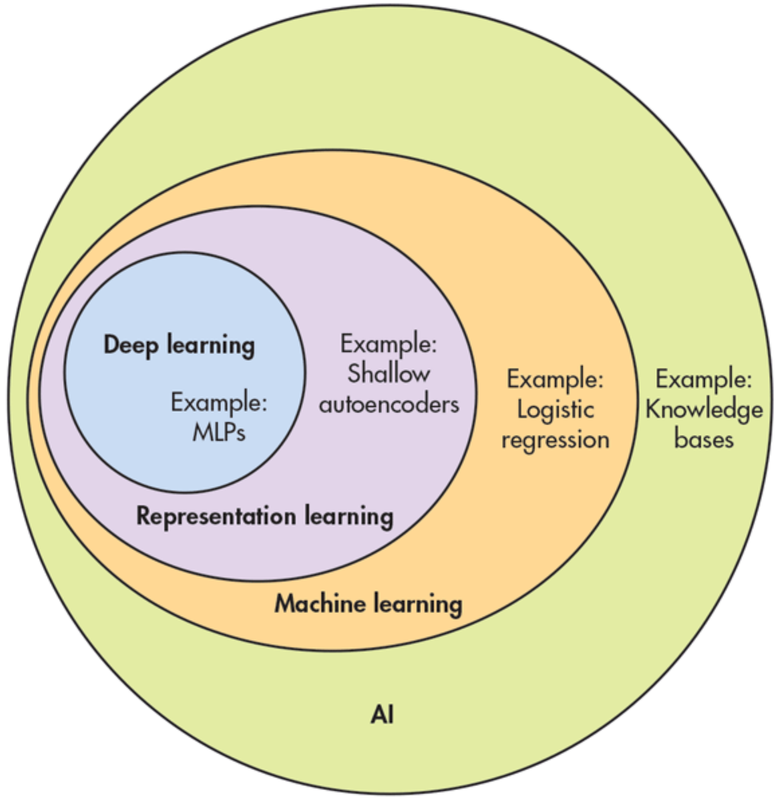
- Arthur Samuel (1959)

"A computer program is said to learn from experience E with some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."

- Tom M. Mitchell (1997)

What is machine learning?





Timeline of Machine Learning

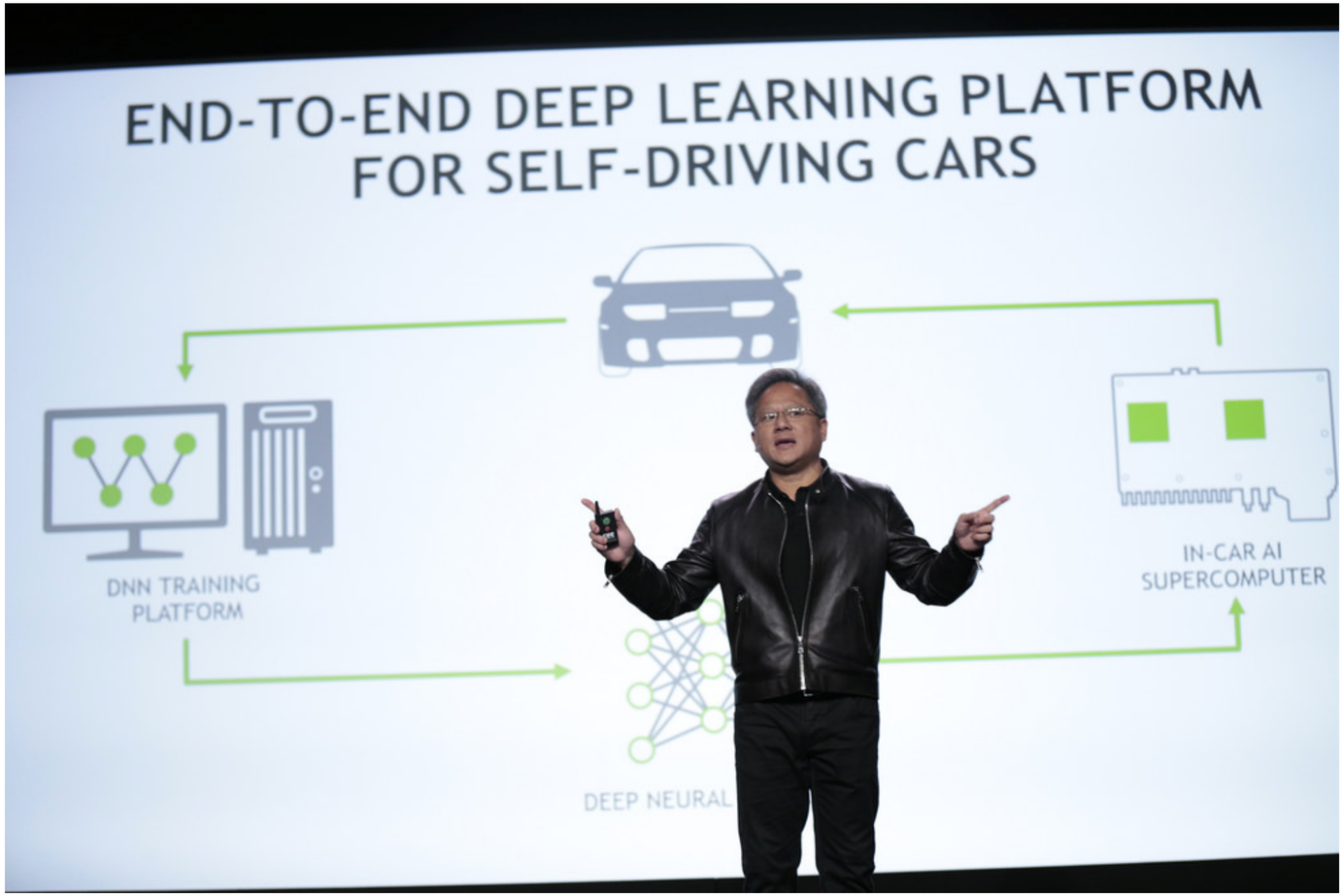
- **1805:** Legendre discovers the least squares method, the earliest form of linear regression.
- **1936:** Fisher proposes linear discriminant analysis.
- **1940s:** Various authors propose logistic regression.
- **1951:** Minsky and Edmonds build the first neural network machine, the SNARC.
- **1957:** Rosenblatt invents the perceptron, a binary classifier.
- **1967:** The nearest neighbor algorithm is created.
- **1970s:** AI winter caused by pessimism about machine learning effectiveness.
- **1980s:** Breiman, Friedman, Olshen, and Stone introduce CARTs. Backpropagation is rediscovered, causing a resurgence in machine learning research.
- **1995:** Ho describes random forests; Cortes and Vapnik introduce SVMs.
- **1997:** IBM's Deep Blue beats Kasparov, the world champion at chess.
- **2009:** ImageNet is created in Fei-Fei Li's group at Stanford. A catalyst for the current AI boom.
- **2016:** Google's AlphaGo defeats an unhandicapped human professional at Go.

Machine Learning Applications

Intelligent personal assistants (Alexa, Google Assistant, Siri)

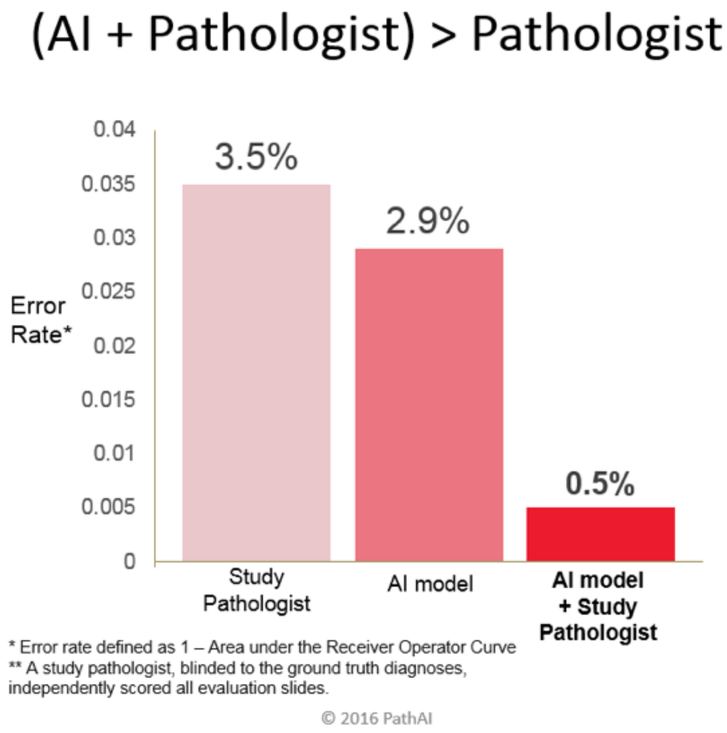
Machine Learning Applications

Self-driving cars



Machine Learning Applications

Deep learning for oncology



Machine Learning Applications

Recommender systems

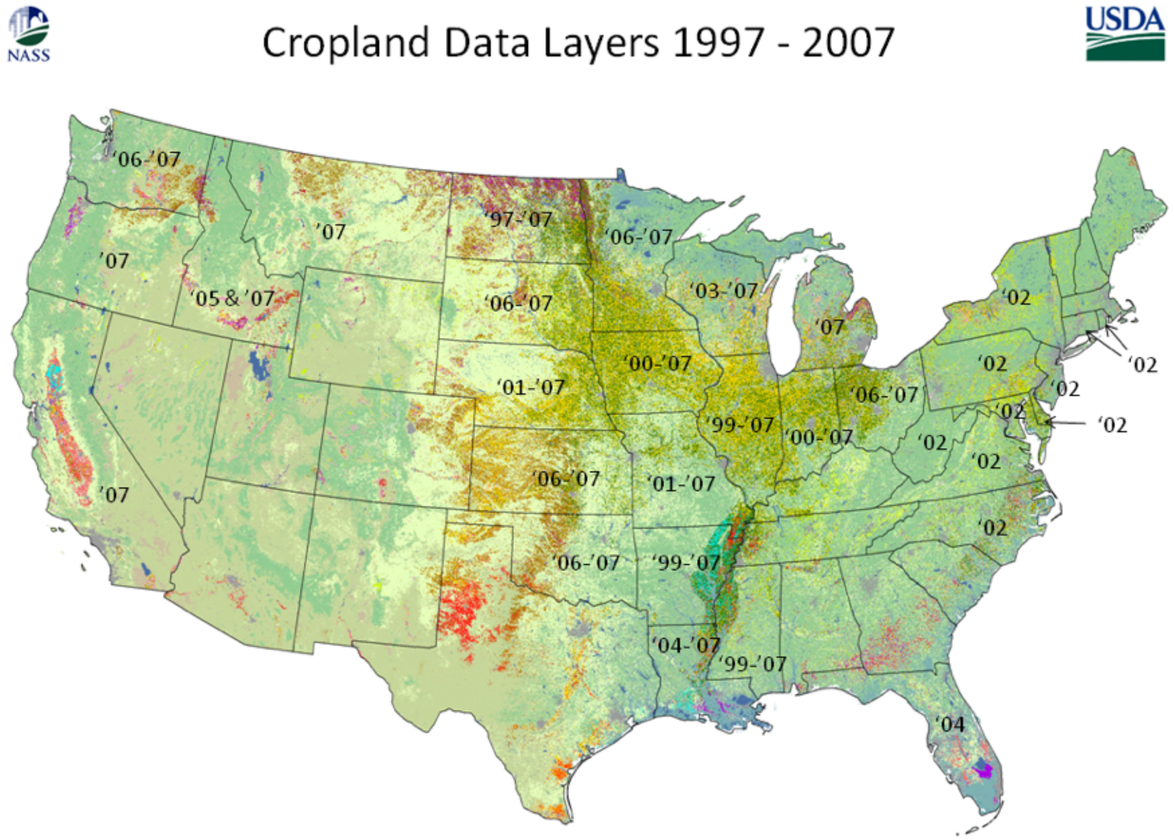
Machine Learning Applications

Legal contract review

LawGeex

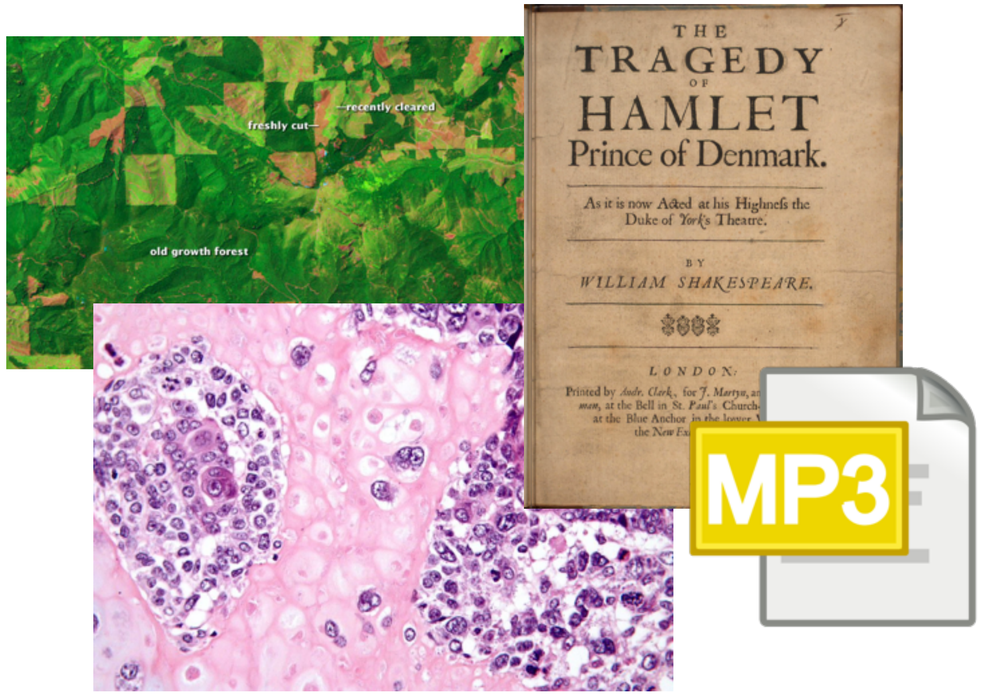
Machine Learning Applications

Mapping agriculture from satellite imagery

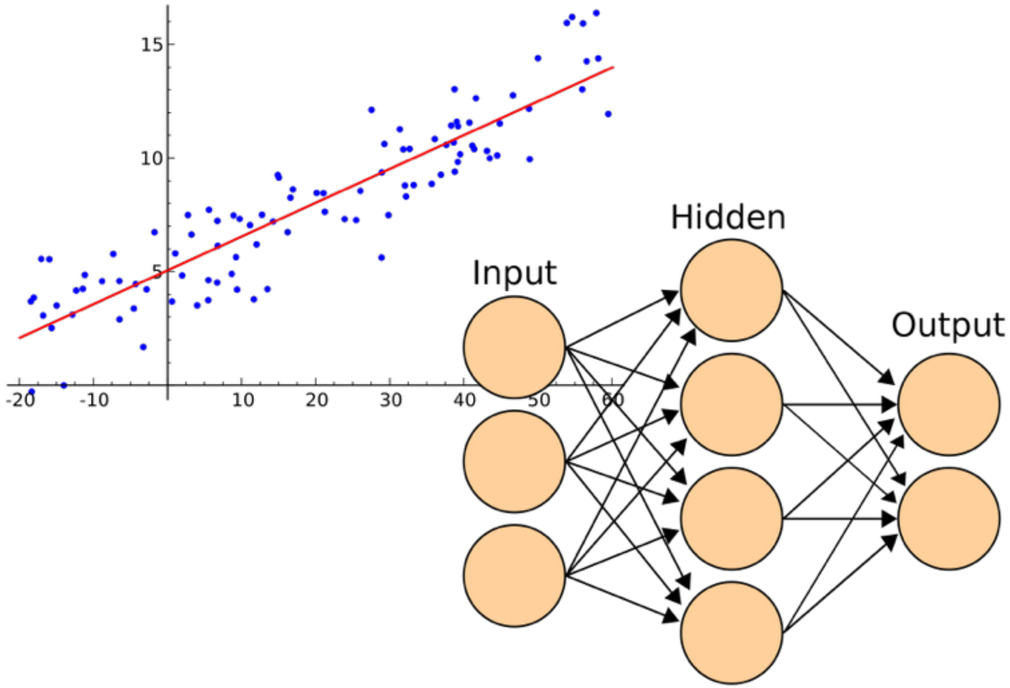


Machine Learning Applications

Data



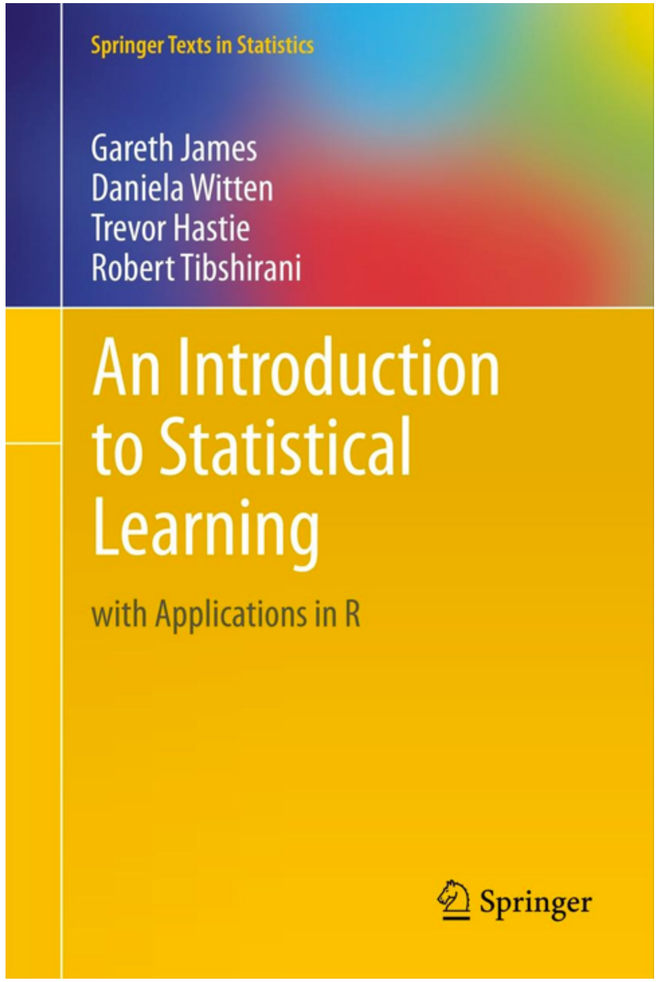
Model



Course Texts

An Introduction to Statistical Learning with Applications in R
by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani

The Elements of Statistical Learning
by Trevor Hastie, Robert Tibshirani, and Jerome Friedman



Machine Learning Overview

Problem Set-up

$\mathbf{X}$: input variables (predictors, independent variables, features)

$\mathbf{Y}$: output variable (response, dependent variable)

Machine learning: estimate a function f that describes the relationship between predictors and response

$$\mathbf{Y} = f(\mathbf{X}) + \varepsilon$$

How to find f?

Training dataset containing n samples $\mathrm{i} = 1, 2, \ldots, n$

Input, output pairs: $(\mathrm{X}^{(1)}, \mathrm{Y}^{(1)}), (\mathrm{X}^{(2)}, \mathrm{Y}^{(2)}), \ldots (\mathrm{X}^{(n)}, \mathrm{Y}^{(n)})$

We will use these observations to build our model f.

$$\mathrm{Y} = f(\mathrm{X}) + \varepsilon$$

Algorithms vary in how they use this data and in the assumptions they place on f.

Prediction vs. Inference

Prediction:

- Predict response $\mathbf{Y}$ given inputs $\mathbf{X}$.

Inference:

- Understand the relationship between $\mathbf{Y}$ and individual predictors $\mathbf{X}_i$.

Do not want a black-box model.

Choosing an ML Algorithm

Which algorithm you use for a task will depend on:

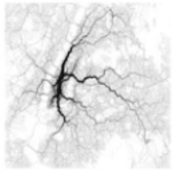
- The type of problem you are trying to solve
- The type of data you have access to

Note that it's possible to have data ill-suited for the problem of interest. In this case, algorithms won't save you.

Machine Learning

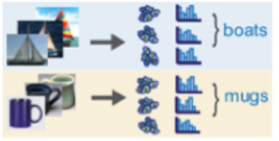
Unsupervised Learning

- Big Data Visualisation



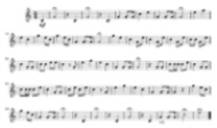
- Structure Discovery

- Dimension Reduction
- Feature Extraction

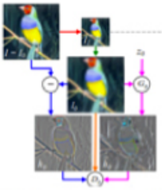


- Generative Networks
  - 2D to 3D Modelling

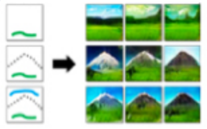
  - Music Generation



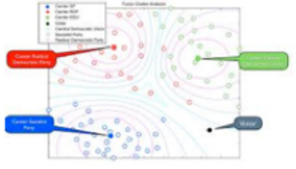
  - Pattern Modelling



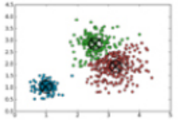
  - Image Generation



- Clustering
  - Recommender Systems



  - Customer Segmentation



  - Targeted Marketing

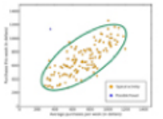
Semi-Supervised Learning

- Text Classification

- Lane Finding

Supervised Learning

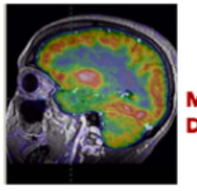
- Fraud Detection



- Gender Detection

- Classification
  - Image Classification

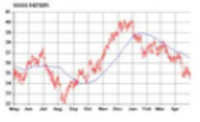
  - Medical Diagnosis



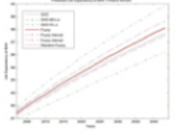
- Regression
  - Weather Forecasting

  - Advertising Reach Prediction

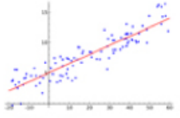
  - Sales Growth Prediction



  - Estimating Life Expectancy



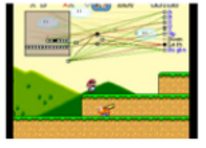
  - Market Forecasting



Reinforcement Learning

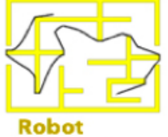
- Real-time Decisions

- Game AI



- Skill Acquisition

- Robot Navigation



- Learning Tasks

By Benjamin Timmermans. http://btimmermans.com

Two Categories of Learning

Supervised learning:

- Builds a statistical model to predict an output from inputs

Unsupervised learning:

- Learns structure from data without supervising output

Supervised Learning

Training data contains both the input variables and the associated response

➡ Mathematically, $\mathbf{X}^{(i)}$ and associated $\mathbf{Y}^{(i)}$ are available to learning algorithm for training

Goal: *generalize* to new data

Unsupervised Learning

Training data contains measurements for each observation, but no associated response of interest

➡ Mathematically, $\mathbf{X}^{(i)}$ are available but $\mathbf{Y}^{(i)}$ are not

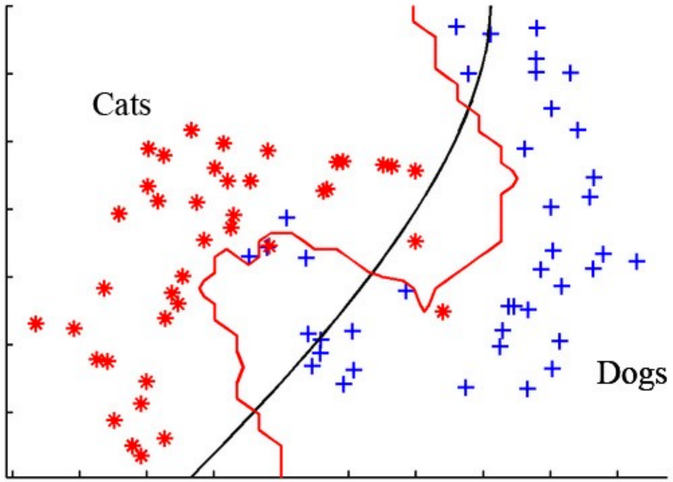
Goal: *understand relationships* between variables or among observations

Two Types of Supervised Learning

Two Types of Supervised Learning

Classification

- Output is qualitative (categorical)
- E.g. predict whether a credit card transaction is fraudulent



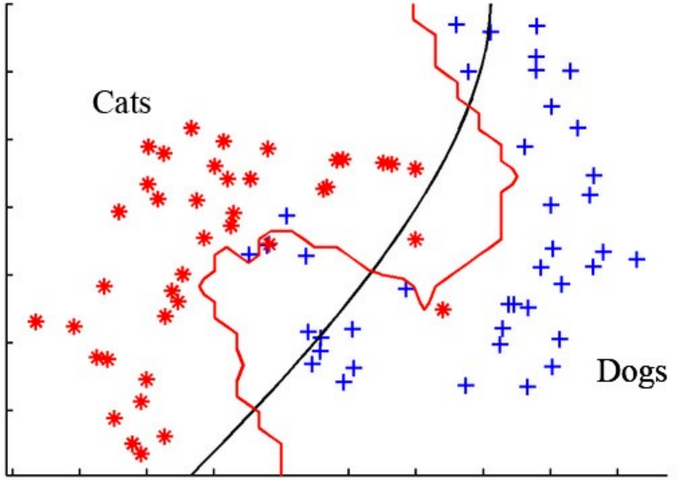
Two Types of Supervised Learning

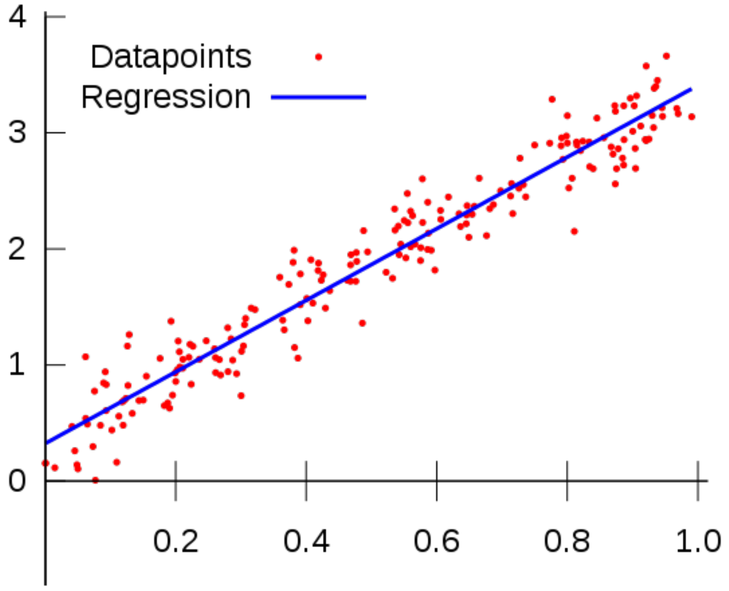
Classification

- Output is qualitative (categorical)
- E.g. predict whether a credit card transaction is fraudulent

Regression

- Output is quantitative (continuous or ordered)
- E.g. predict the value of a stock tomorrow





Classification and Regression

Classification can often be formulated as a regression problem

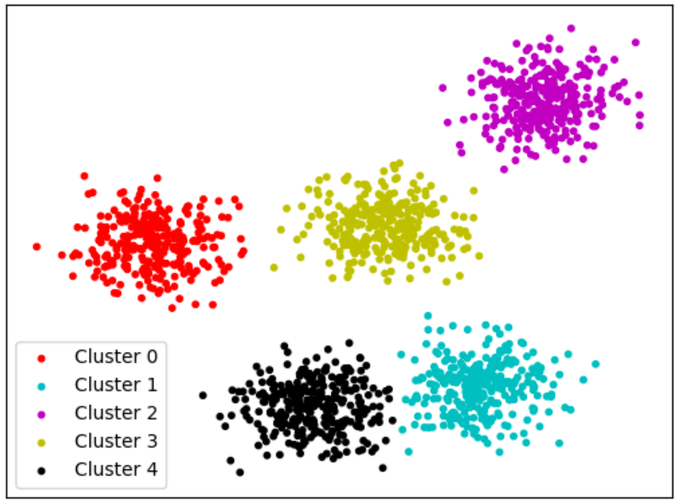
- For a two-class (binary) problem: “What is the probability that observation belongs to class 1?” Probability lies in [0,1]
- Some methods work well on both types of problems (e.g. neural networks)

Two Types of Unsupervised Learning

Two Types of Unsupervised Learning

Clustering

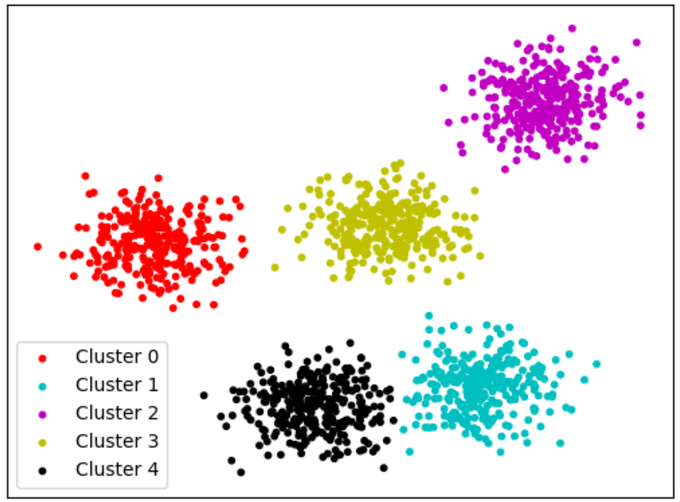
- Partition data into subsets that share common characteristics



Two Types of Unsupervised Learning

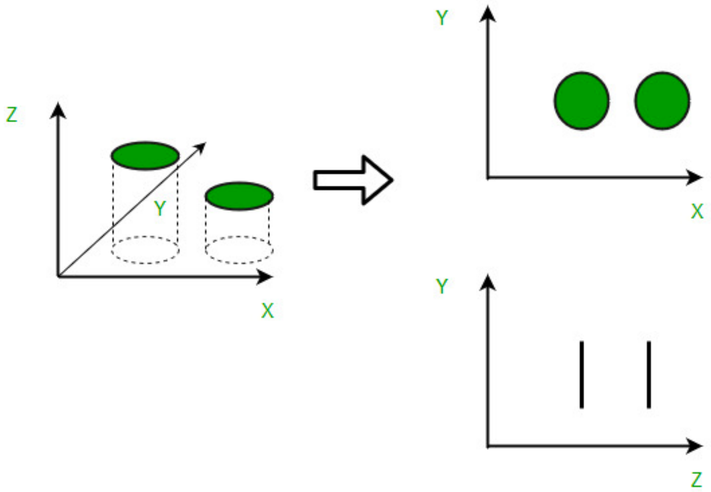
Clustering

- Partition data into subsets that share common characteristics

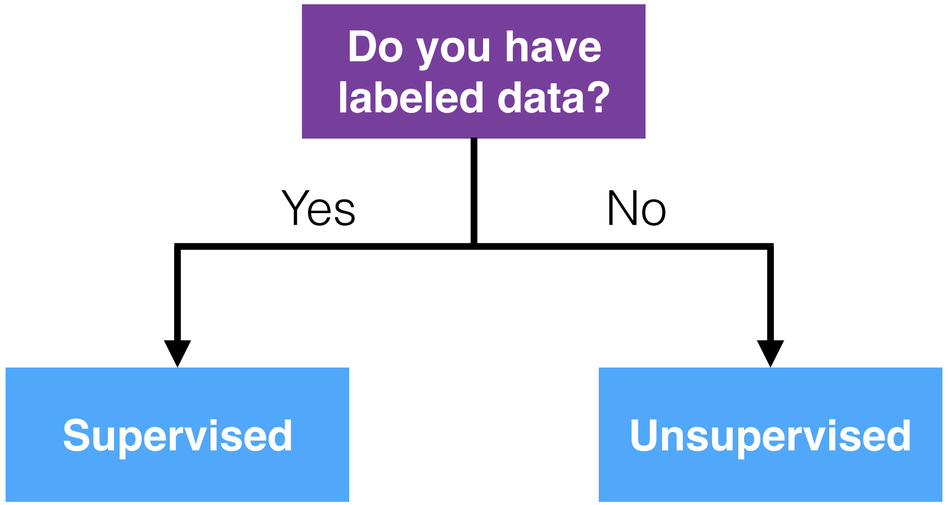


Dimensionality reduction

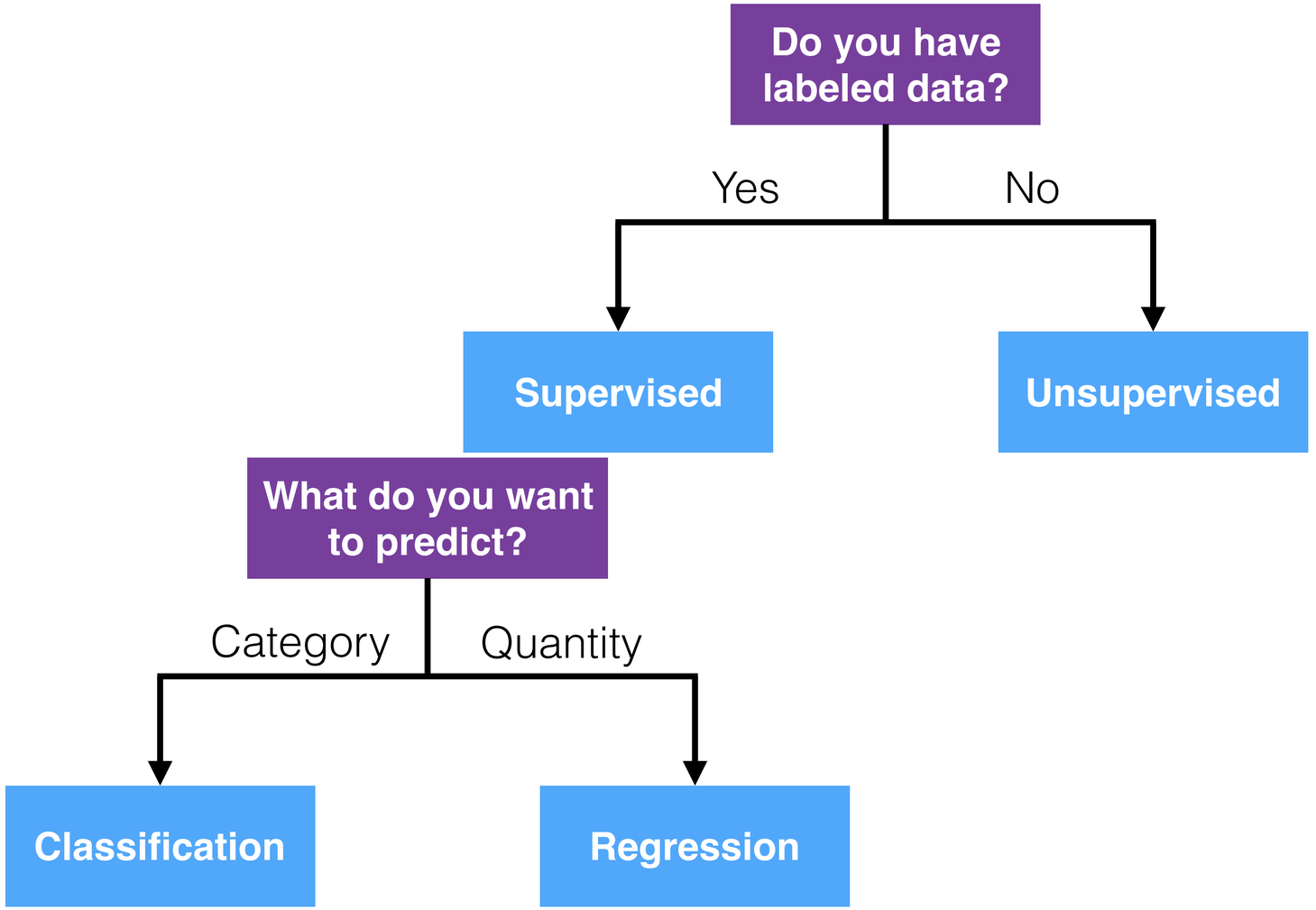
- Create new features from original inputs that retain important information



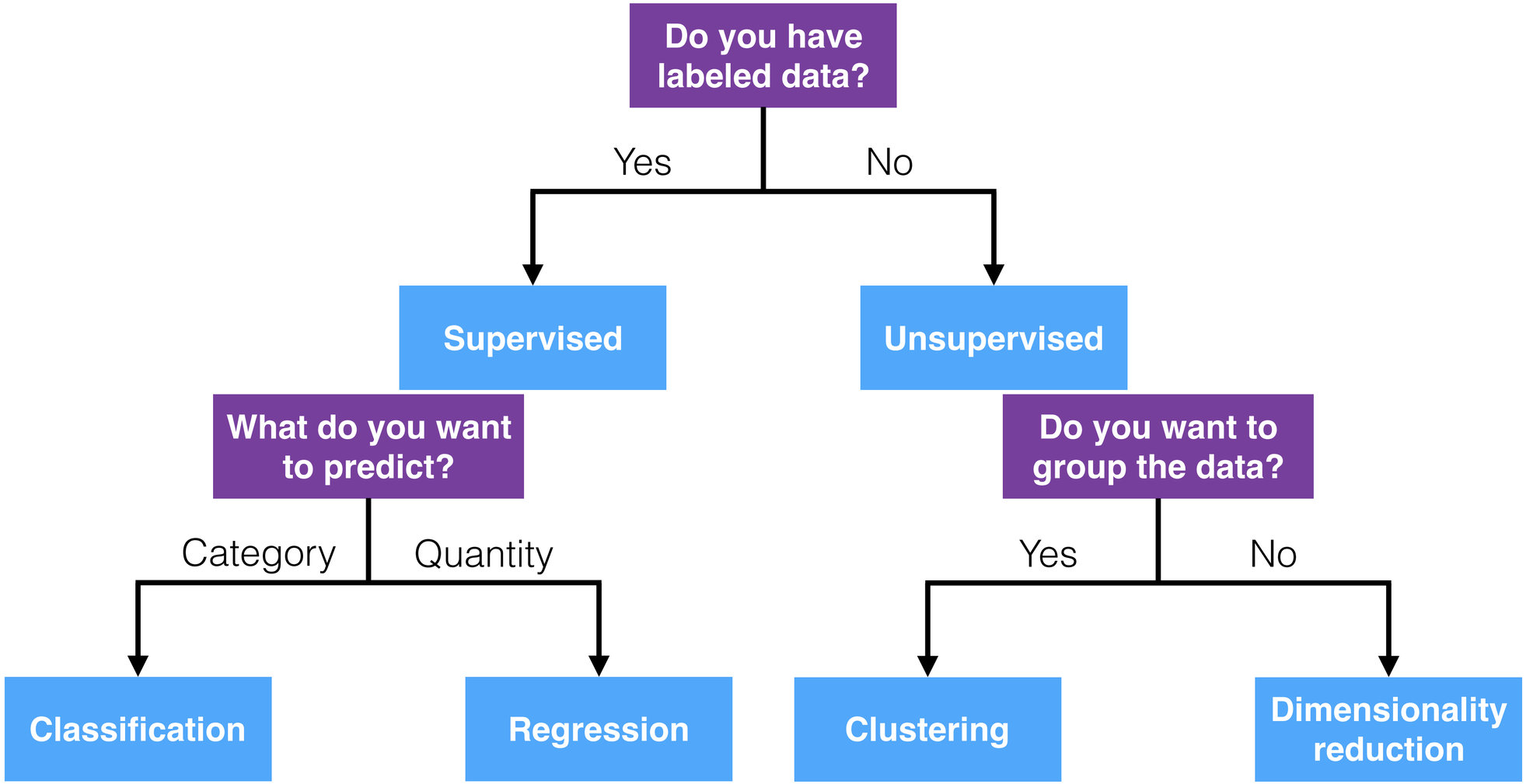
Choosing an ML Algorithm

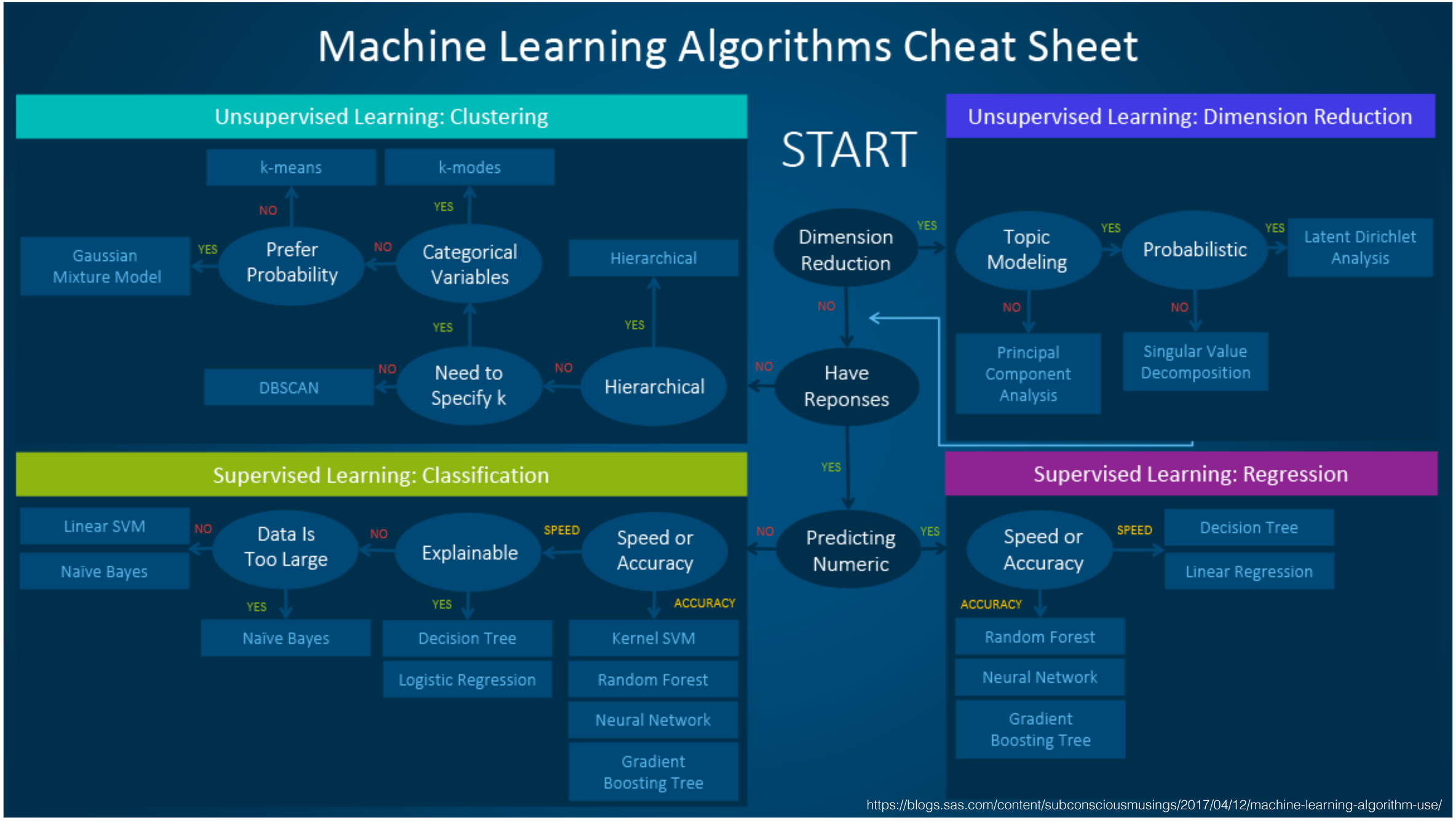


Choosing an ML Algorithm

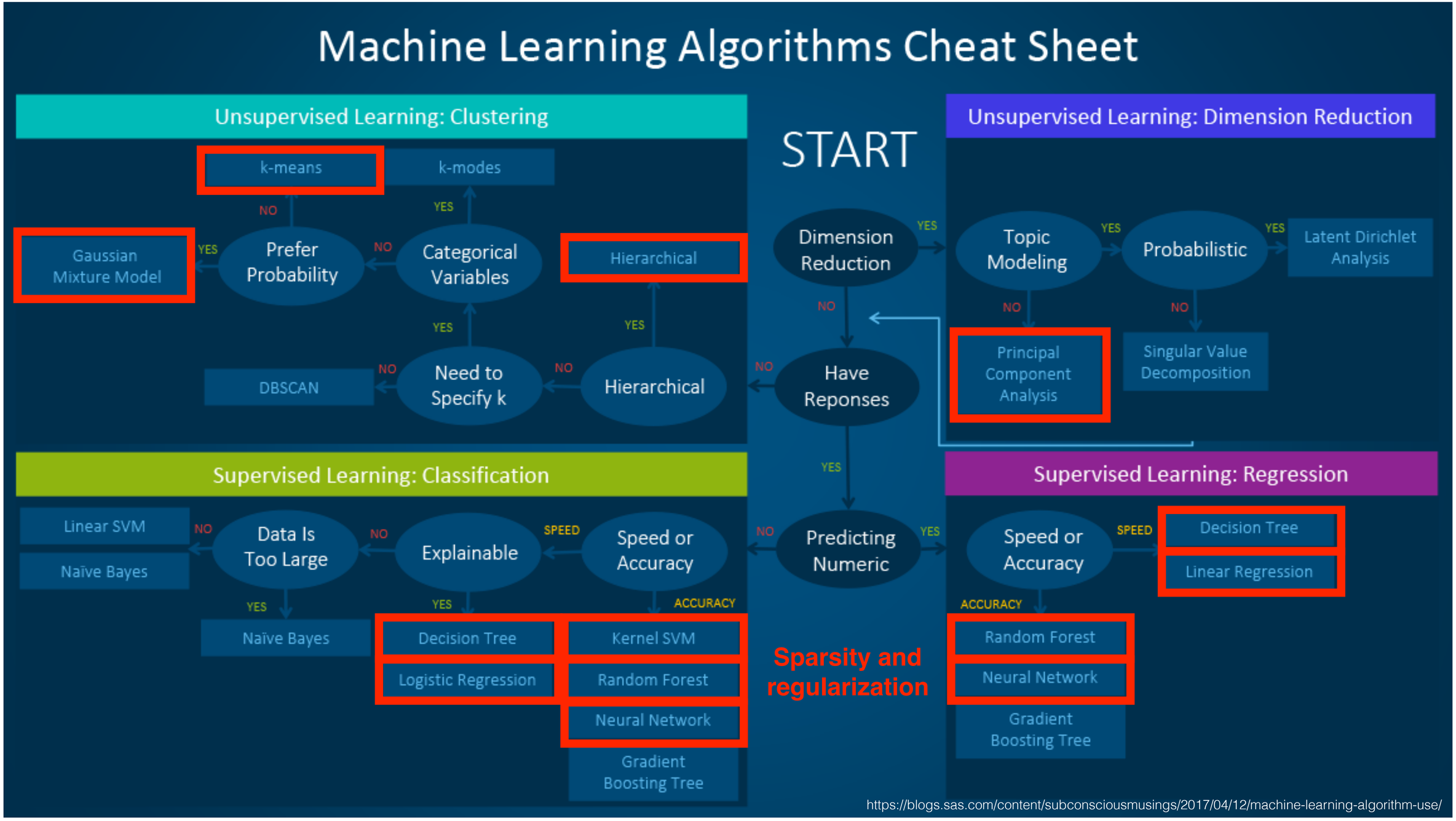


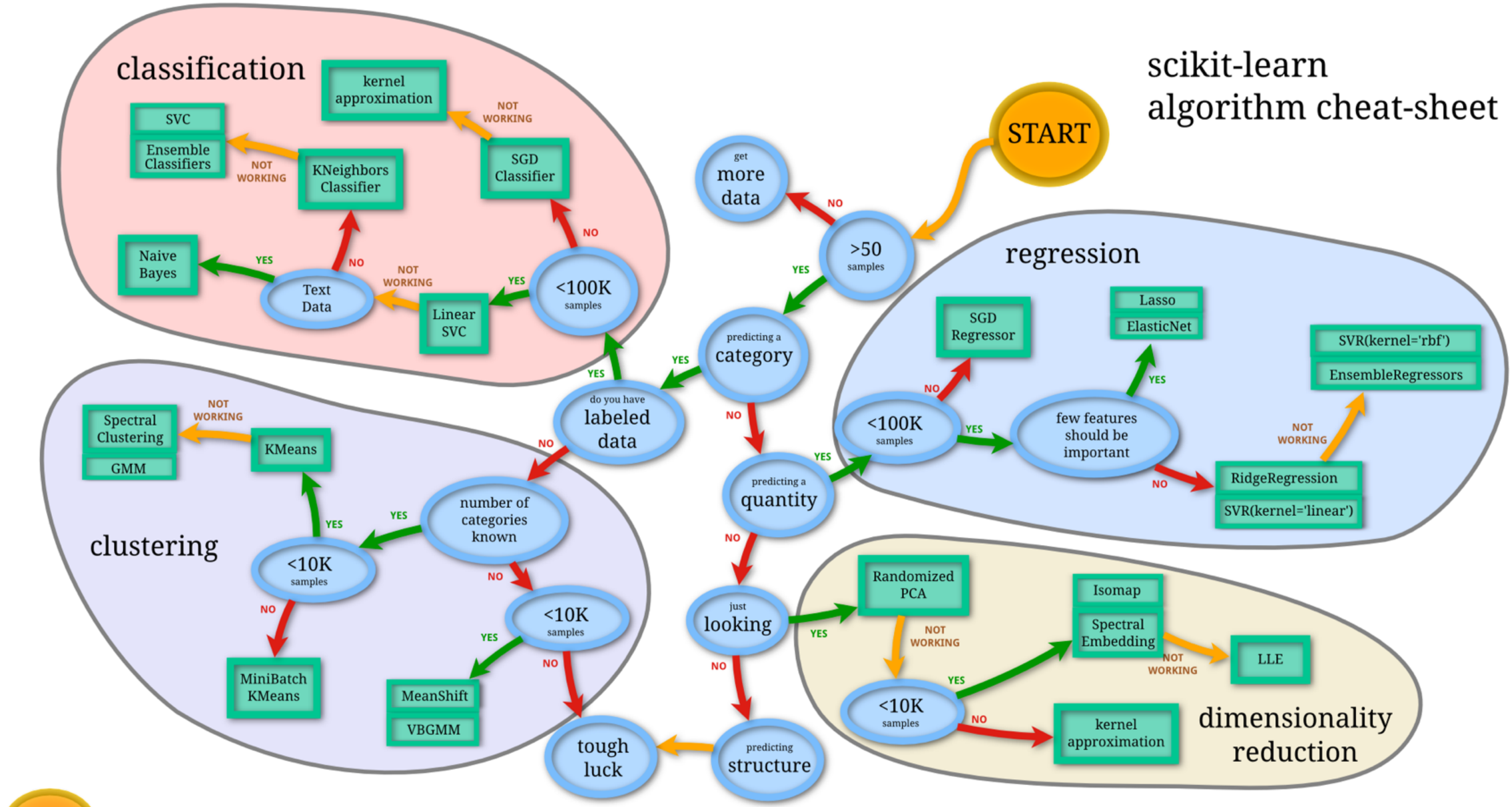
Choosing an ML Algorithm



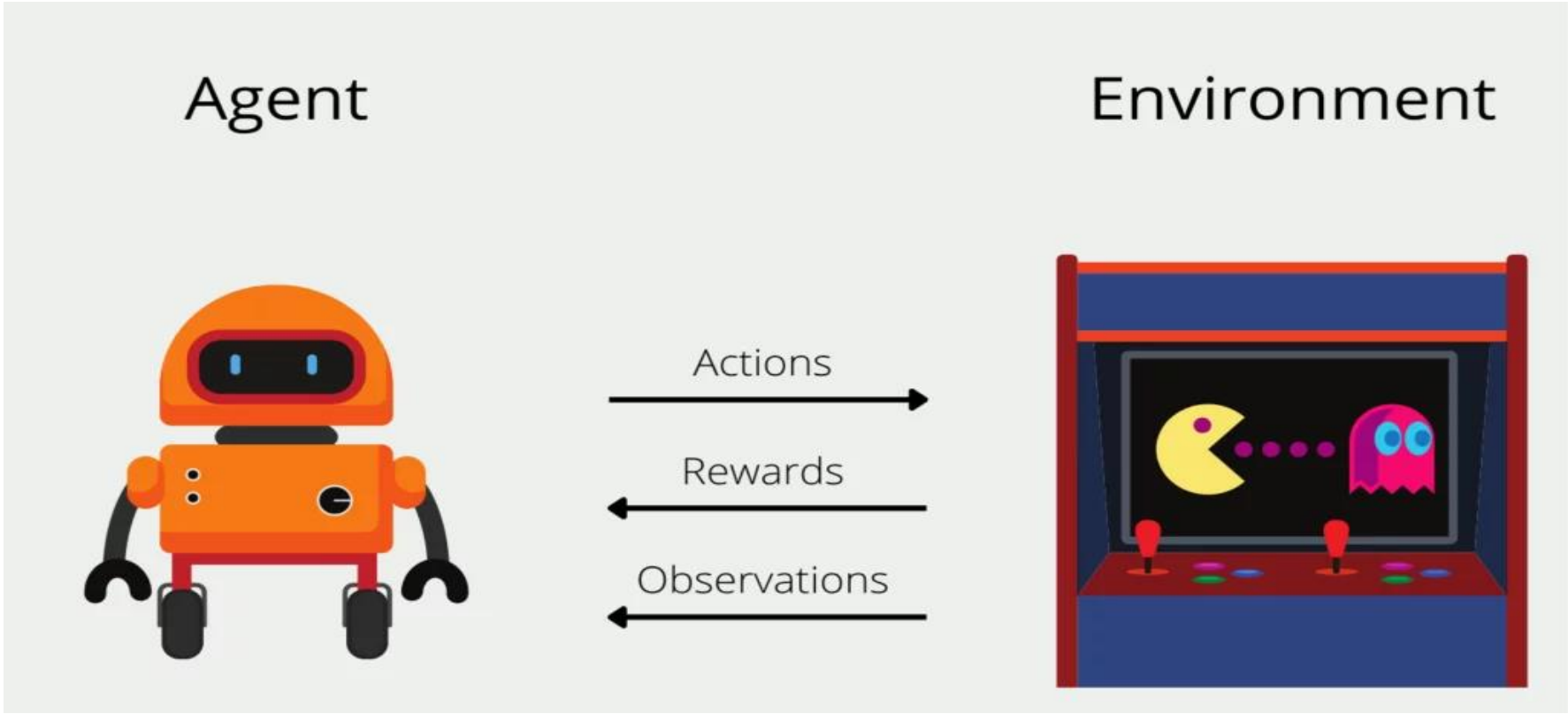


https://blogs.sas.com/content/subconsciousmusings/2017/04/12/machine-learning-algorithm-use/


Machine Learning Algorithms Cheat Sheet
Unsupervised Learning: Clustering
START
Unsupervised Learning: Dimension Reduction
k-means
k-modes
NO
YES
Gaussian Mixture Model
YES
Prefer Probability
NO
Categorical Variables
Hierarchical
YES
YES
NO
DBSCAN
NO
Need to Specify k
NO
Hierarchical
Dimension Reduction
YES
Topic Modeling
YES
Probabilistic
YES
Latent Dirichlet Analysis
NO
NO
Principal Component Analysis
Singular Value Decomposition
NO
Have Reponses
YES
Supervised Learning: Classification
Supervised Learning: Regression
Linear SVM
NO
Data Is Too Large
NO
Explainable
SPEED
Speed or Accuracy
NO
Predicting Numeric
YES
Speed or Accuracy
SPEED
Decision Tree
Naïve Bayes
Linear Regression
YES
YES
ACCURACY
ACCURACY
Naïve Bayes
Decision Tree
Kernel SVM
Random Forest
Logistic Regression
Neural Network
Gradient Boosting Tree
Sparsity and regularization
Random Forest
Neural Network
Gradient Boosting Tree
https://blogs.sas.com/content/subconsciousmusings/2017/04/12/machine-learning-algorithm-use/


scikit-learn
algorithm cheat-sheet
START
get more data
NO
>50 samples
YES
predicting a category
YES
do you have labeled data
YES
<100K samples
NO
Text Data
YES
Naive Bayes
NO
KNeighbors Classifier
NOT WORKING
SVC
Ensemble Classifiers
NOT WORKING
kernel approximation
NOT WORKING
SGD Classifier
NO
NOT WORKING
Linear SVC
YES
classification
NO
predicting a quantity
YES
<100K samples
NO
SGD Regressor
YES
few features should be important
YES
Lasso
ElasticNet
NO
RidgeRegression
NOT WORKING
SVR(kernel='rbf')
EnsembleRegressors
SVR(kernel='linear')
regression
NO
number of categories known
YES
<10K samples
NO
KMeans
YES
Spectral Clustering
NOT WORKING
GMM
MiniBatch KMeans
NO
<10K samples
YES
MeanShift
VBGMM
NO
clustering
NO
just looking
YES
Randomized PCA
NOT WORKING
<10K samples
YES
Isomap
Spectral Embedding
NOT WORKING
LLE
NO
kernel approximation
dimensionality reduction
NO
predicting structure
tough luck
Back
scikit learn

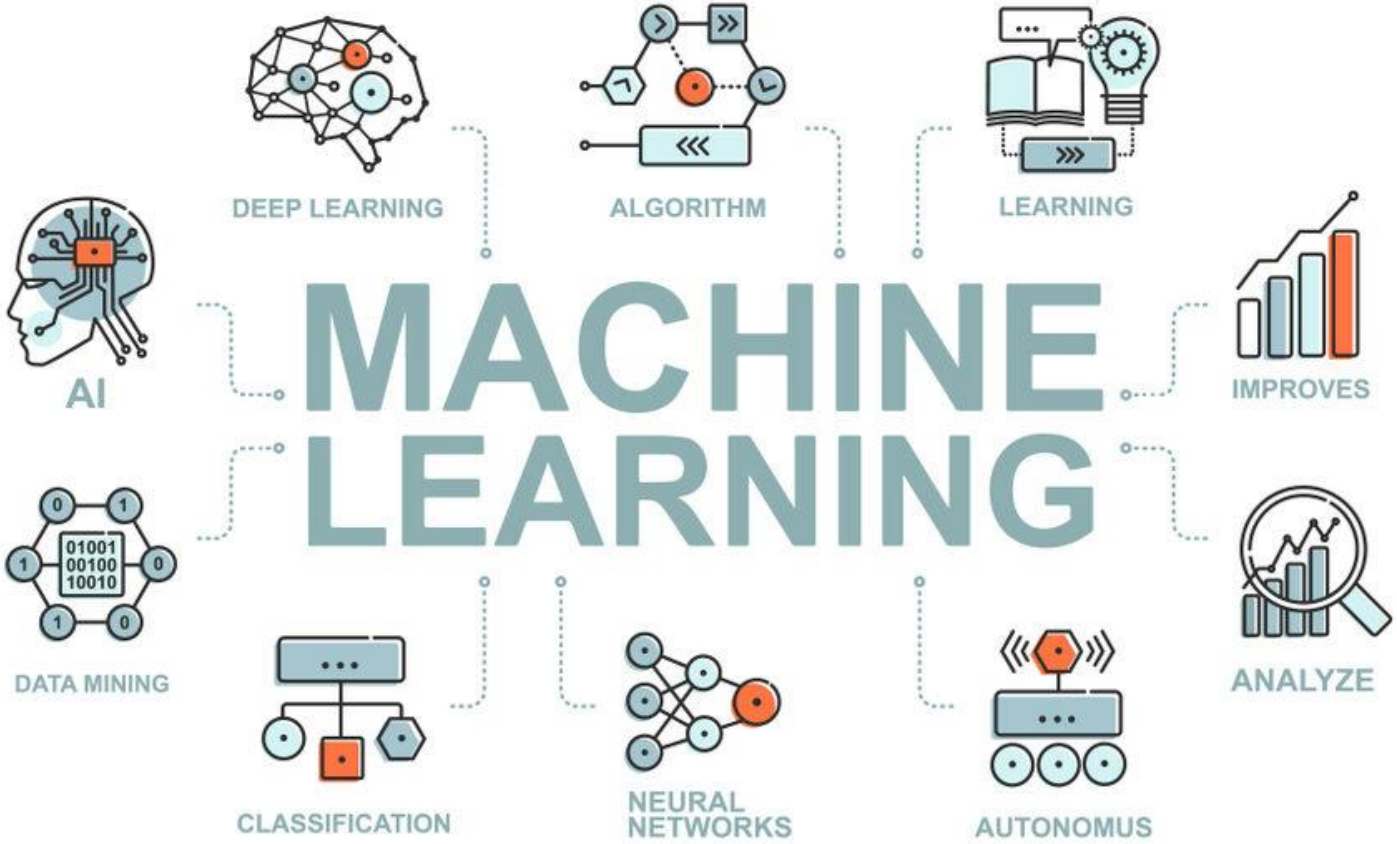
Reinforcement learning



REINFORCEMENT LEARNING ALGORITHMS

Supervised vs Unsupervised vs Reinforcement Learning

Criteria	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Definition	Learns from labeled data	Learns from unlabeled data without guidance	Learns by interacting with the environment and receiving feedback
Type of Problems	Regression, Classification	Clustering, Association	Reward-based decision problems
Type of Data	Labeled data	Unlabeled data	No predefined dataset
Training	Requires external supervision	No supervision needed	No supervision; learns via trial and error
Approach	Maps input to known output	Discovers hidden patterns and relationships	Learns optimal actions via reward signals
Popular Algorithms	Linear Regression, Logistic Regression, SVM, KNN	K-Means, Hierarchical Clustering, PCA	Q-Learning, DQN, SARSA


MACHINE LEARNING
DEEP LEARNING
ALGORITHM
LEARNING
AI
IMPROVES
01001
00100
10010
DATA MINING
ANALYZE
CLASSIFICATION
NEURAL NETWORKS
AUTONOMUS