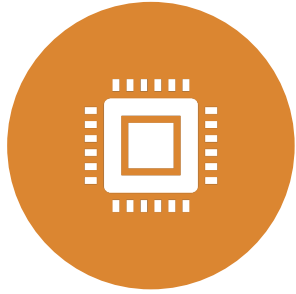


A PRIMER TO MACHINE LEARNING

AIST2010 Lecture 9



What is Machine Learning?



Supervised and Unsupervised Learning

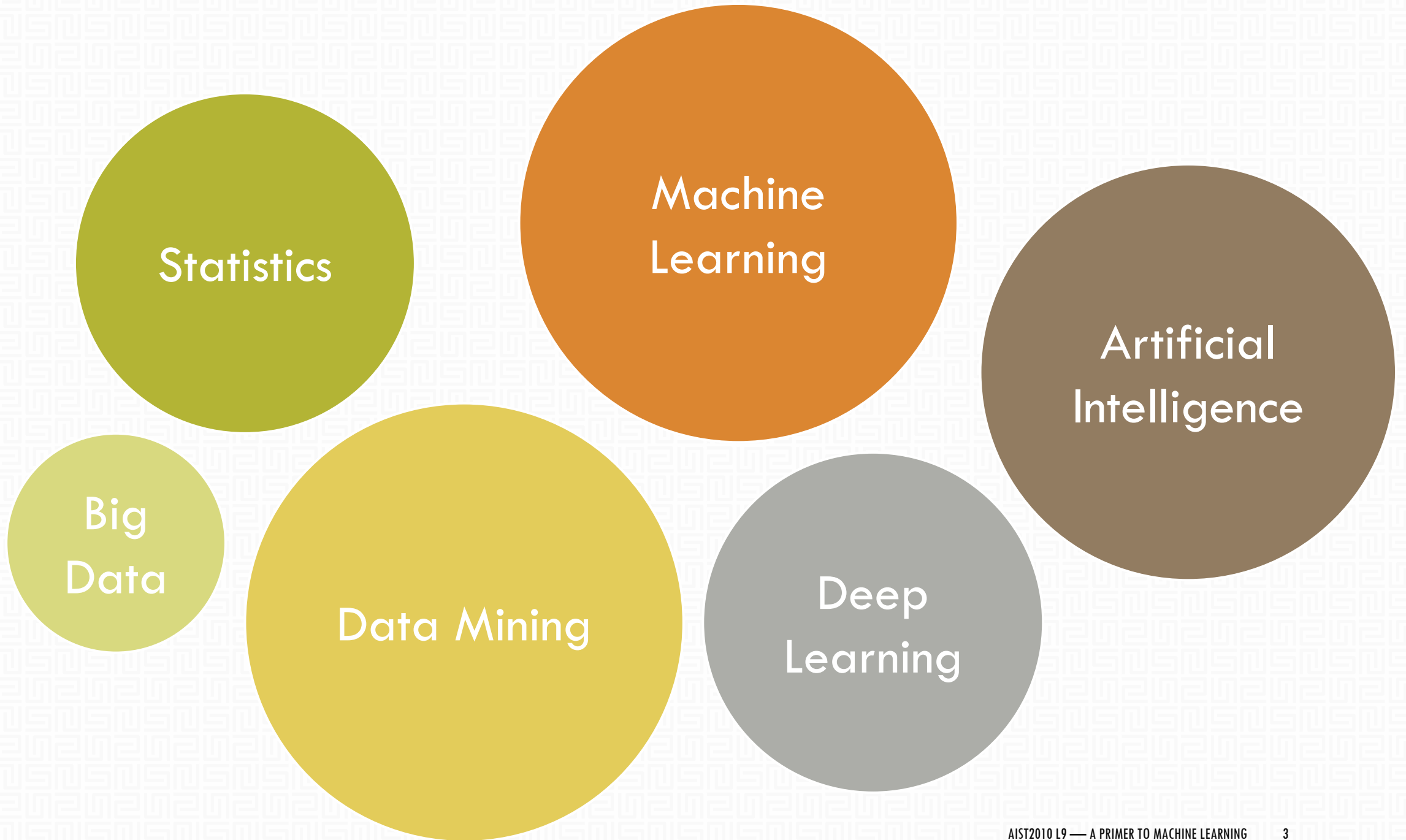


Reinforcement Learning



Artificial Neural Networks

OUTLINE



MACHINE LEARNING

Supervised learning — predicting using trained characteristics

- Classification
- Regression

Unsupervised learning — looking for similarities or patterns

- Clustering
- Dimension reduction

Reinforcement learning — maximizing rewards

MACHINE LEARNING

All the Buzzwords together!

Read: Machine Learning for Everyone

- https://vas3k.com/blog/machine_learning

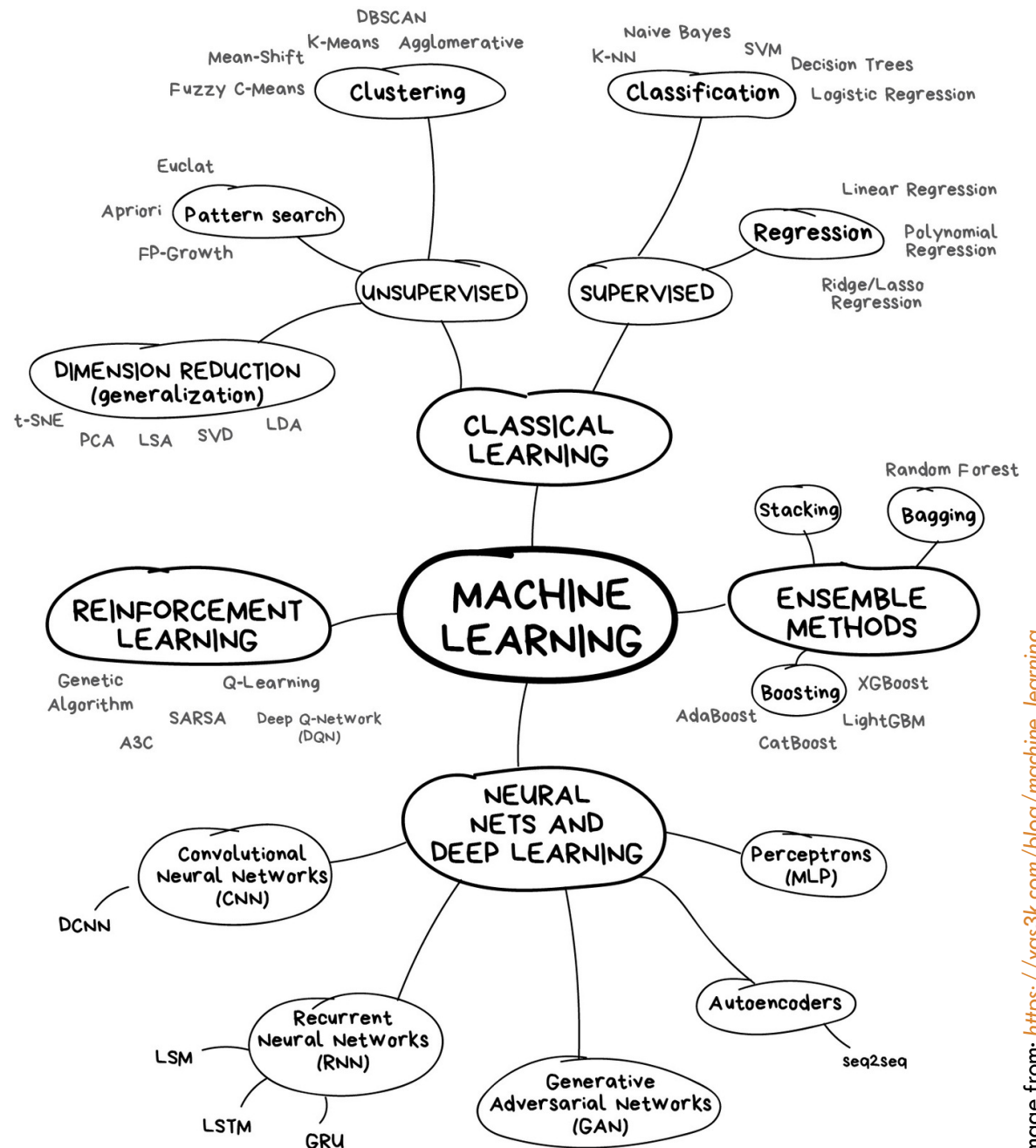
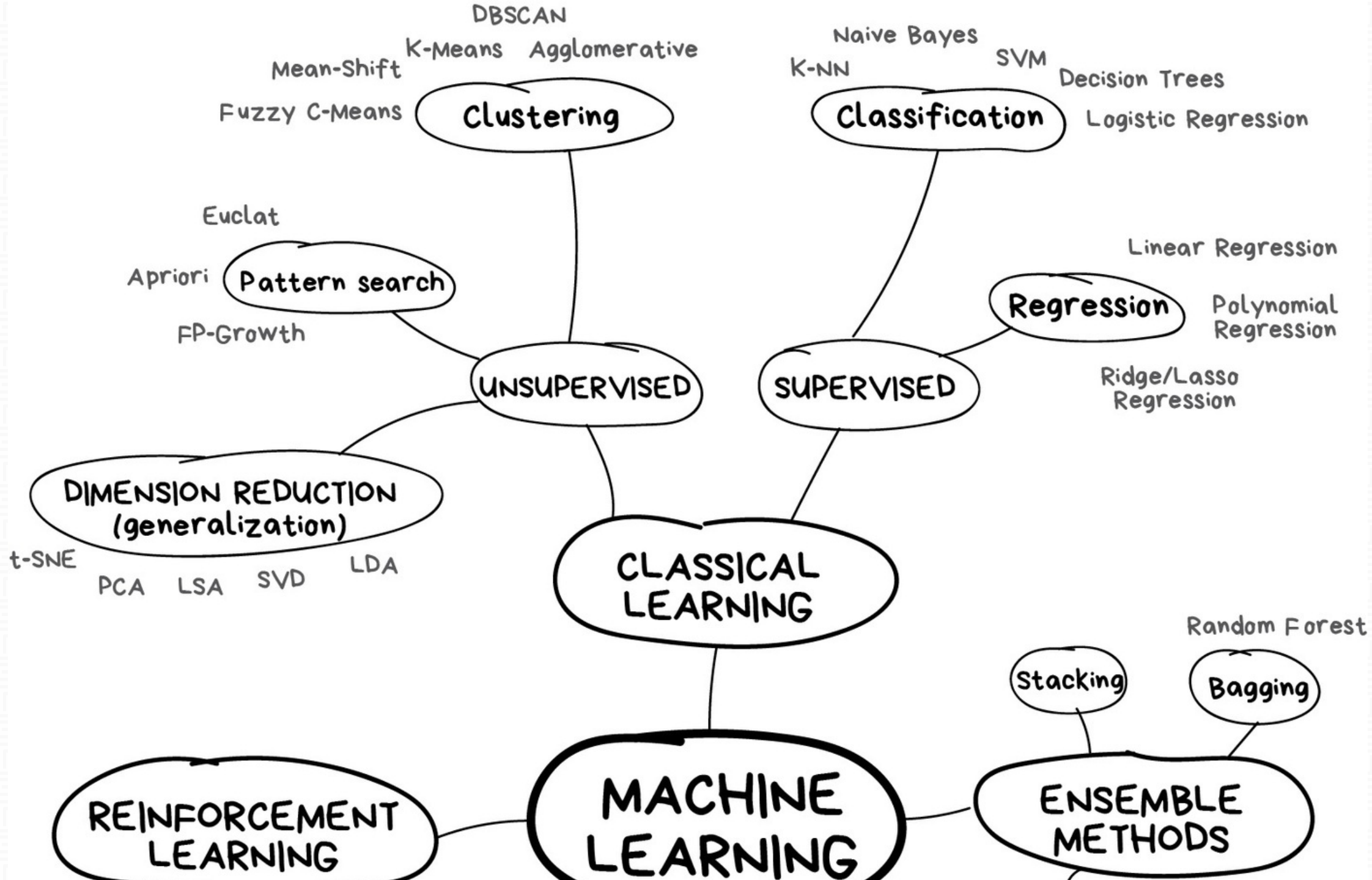
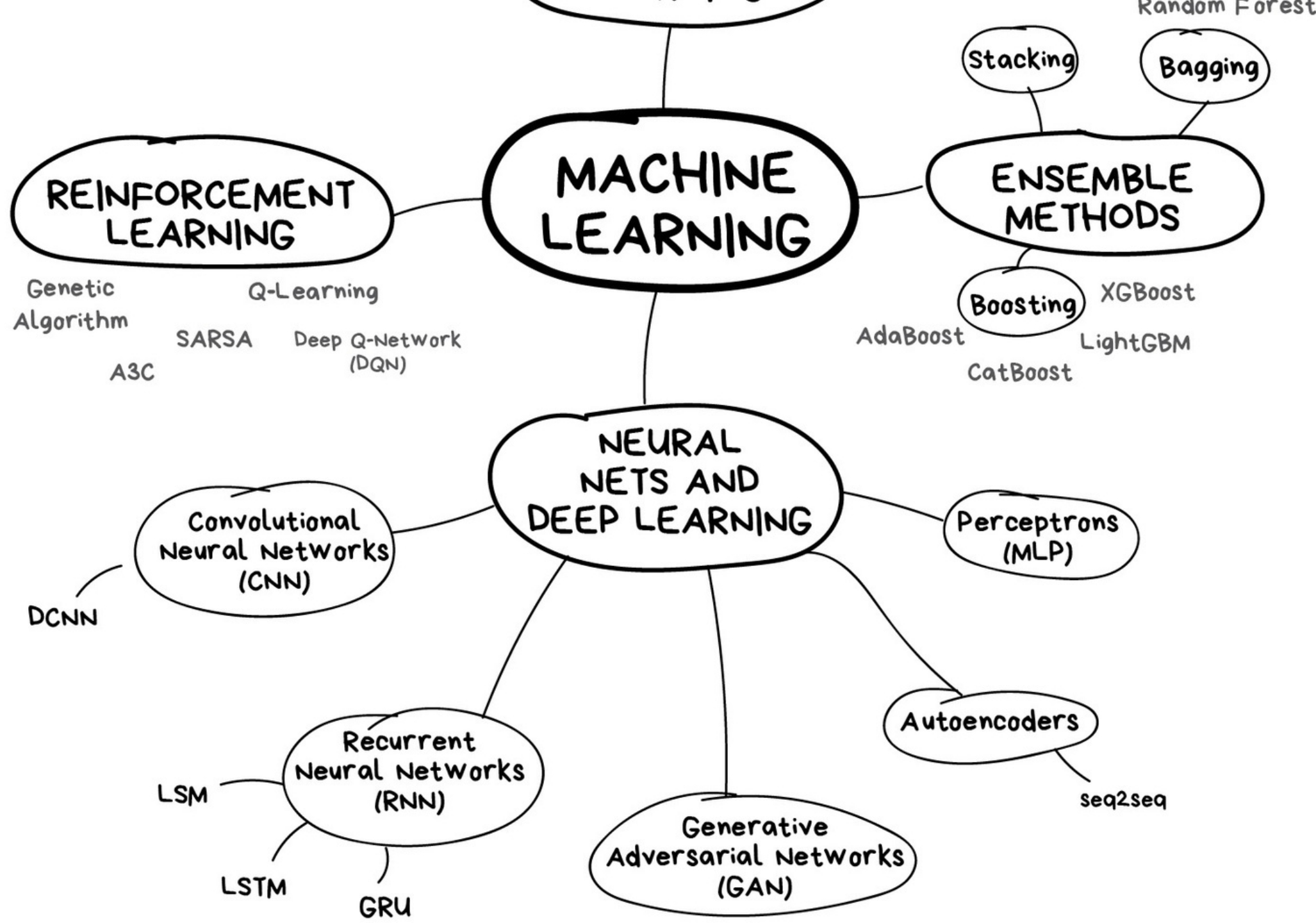
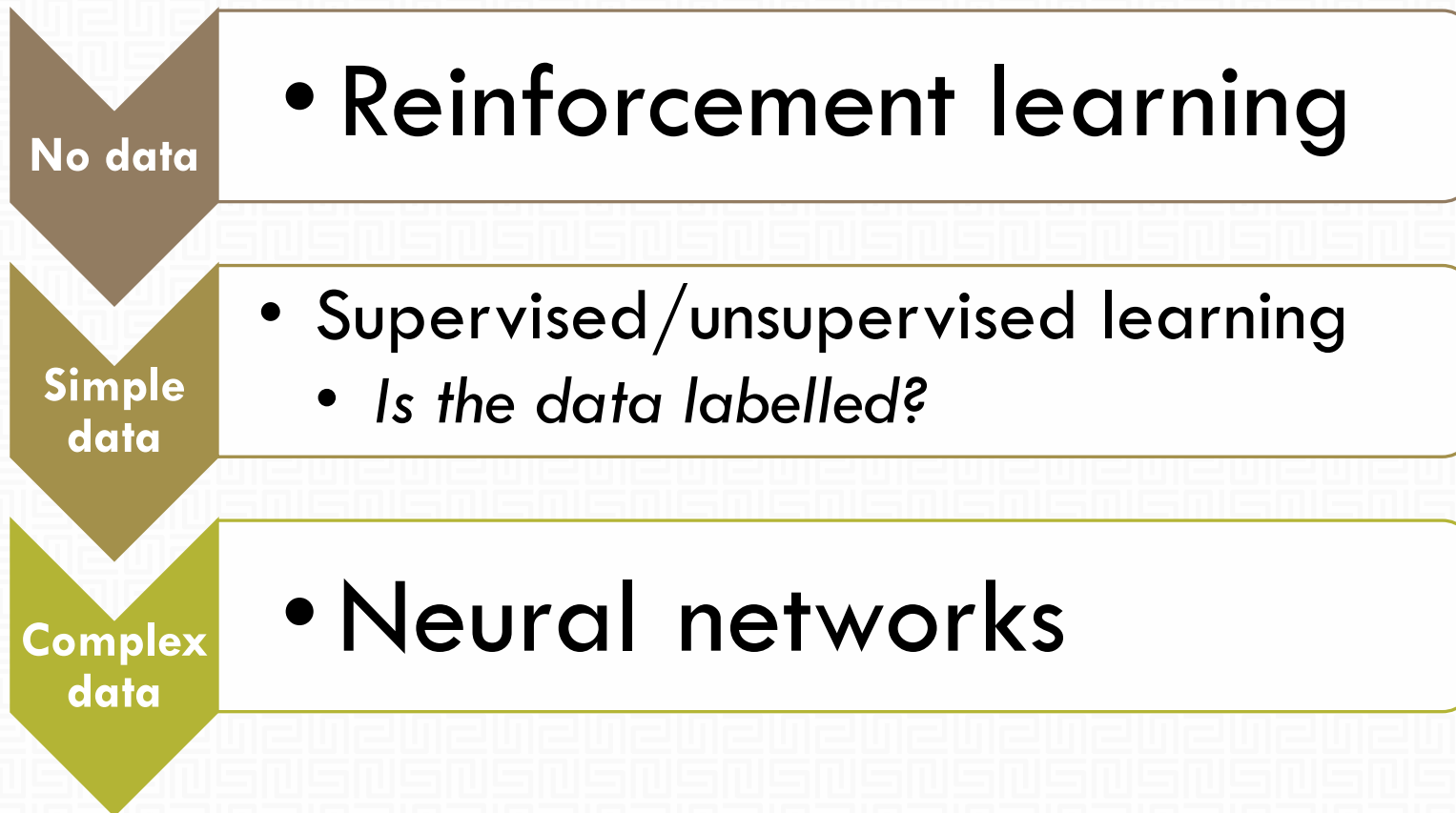


Image from: https://vas3k.com/blog/machine_learning





DEPENDING ON YOUR DATA...



SUPERVISED LEARNING

Labelled data for training

- Learning with example input-output pairs

Accuracy depends on

- The choice of training set
- Feature representation
- Heterogeneity of data
- Redundancy in data

Steps:

1. *Gather training set with appropriate features*
2. *Run algorithm on training set*
3. *Optimize parameters with validation subset (repeatedly)*

Problem of overfitting — working well on training set only

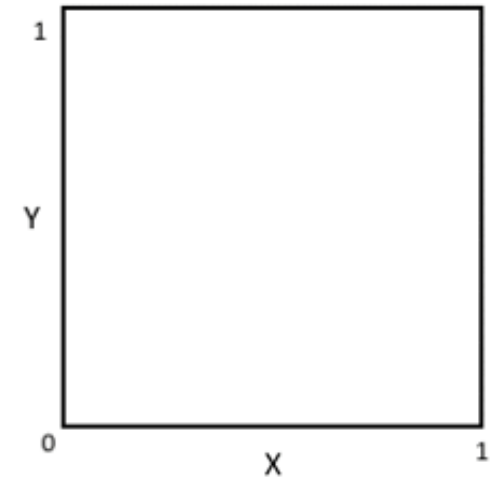
DECISION TREE

Binary questions: Yes vs. No

- Can be probabilities as well

Drawback: easily affected by noise in training data

Enhanced by multiple trees (forests) of subsets of training data



For more tutorials: annalysin.wordpress.com

Image from: <https://annalysin.wordpress.com/2016/07/27/decision-trees-tutorial/>

SUPPORT VECTOR MACHINE (SVM)

Binary separation: where is the maximum separation?

- “Maximum-margin hyperplane”
- “Kernels” to transform polynomial or gaussian separator to linear

Classification or regression analysis

Drawback: only available for 2 classes

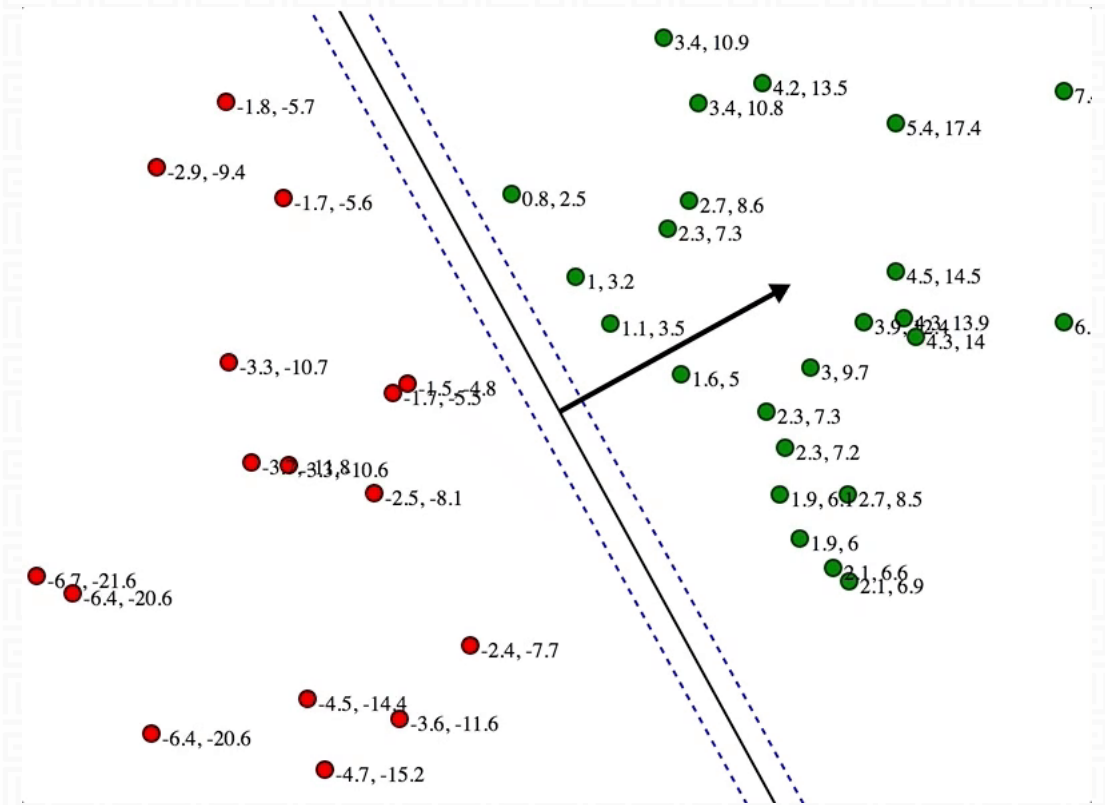


Image from: <https://jeremykun.com/2017/06/05/formulating-the-support-vector-machine-optimization-problem/>

UNSUPERVISED LEARNING

No labels given!

- Algorithms discover pattern and similarities

Clustering

- Summarising data
- Association mining

Dimension reduction: Latent variable models

- Determining extra dimensions/features for supervised learning

K-MEANS CLUSTERING

Partitioning n observations into k clusters

1. Assign (random) cluster centres
2. Calculate distance means to the cluster centres
3. Find new centres to minimize distance (and repeat from step 2)

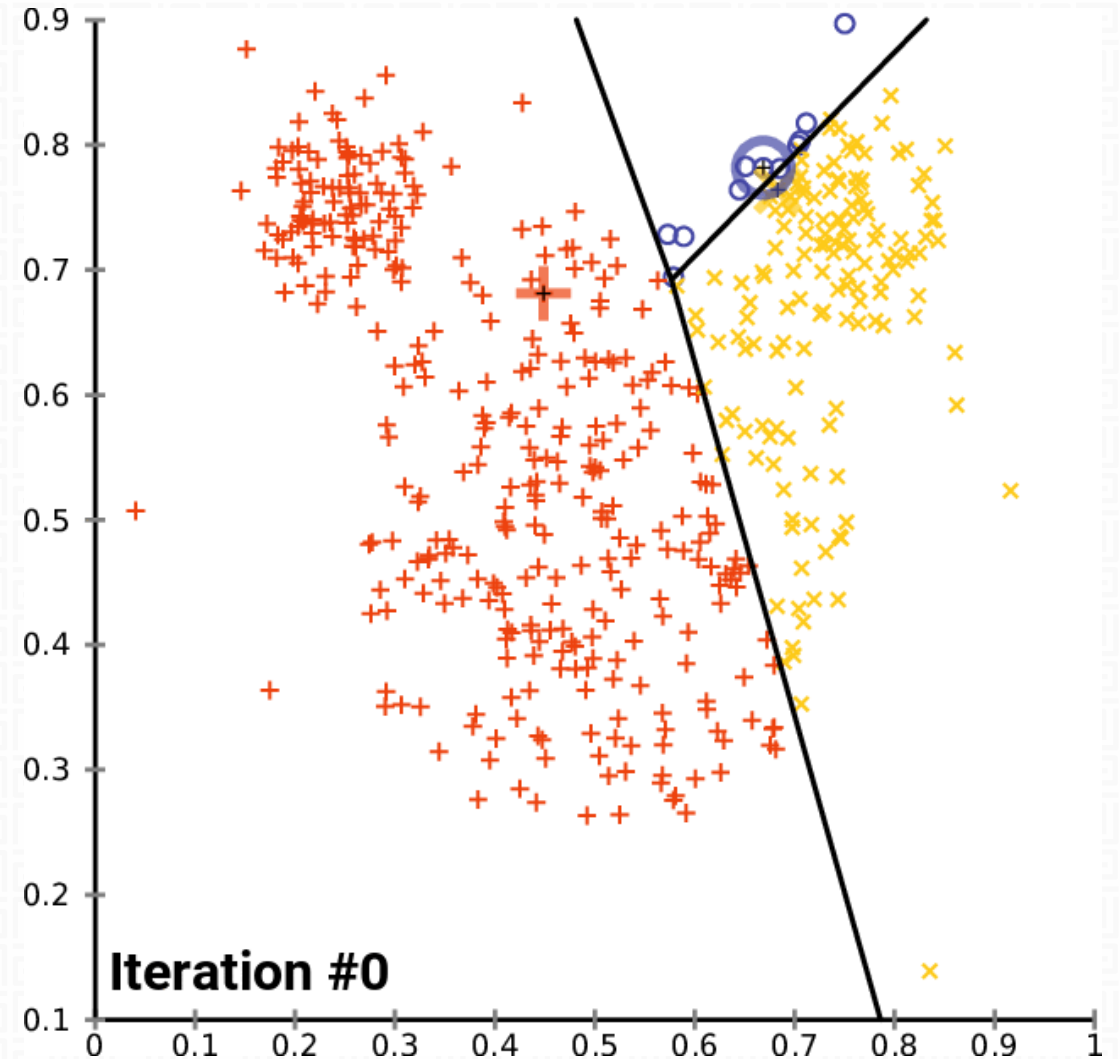


Image from: https://en.wikipedia.org/wiki/K-means_clustering

PRINCIPLE COMPONENT ANALYSIS (PCA)

Transformation to another coordinate system

- Ensure largest possible variance
- Bring out strong patterns in a dataset

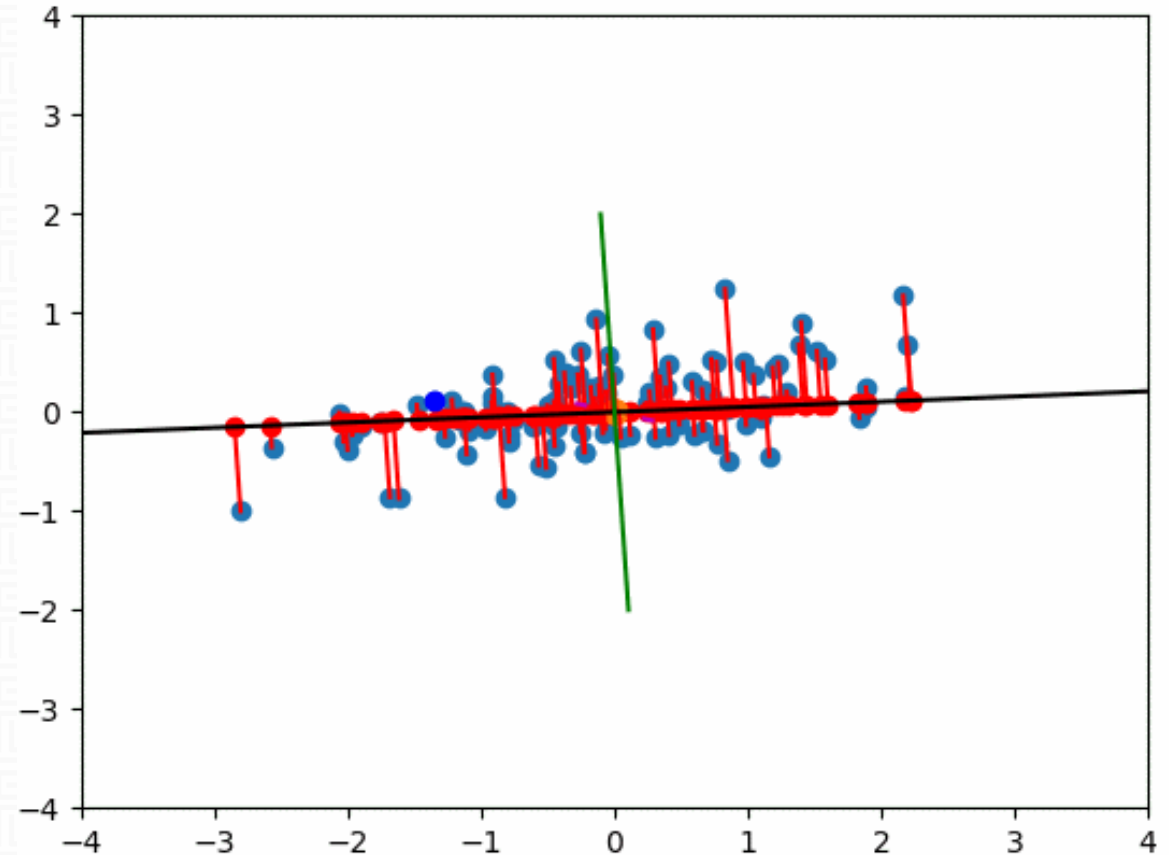


Image from: <https://medium.com/x8-the-ai-community/principal-component-analysis-a-brief-introduction-dc8cf3e03c71>

REINFORCEMENT LEARNING

No training data

- Goal is to maximize the rewards

e.g. *Genetic Algorithm (GA)*

“Survival of the Fittest”

1. Initial random population
2. Evaluate fitness (repeat)
 3. Selection
 4. Cross-over
 5. Mutation

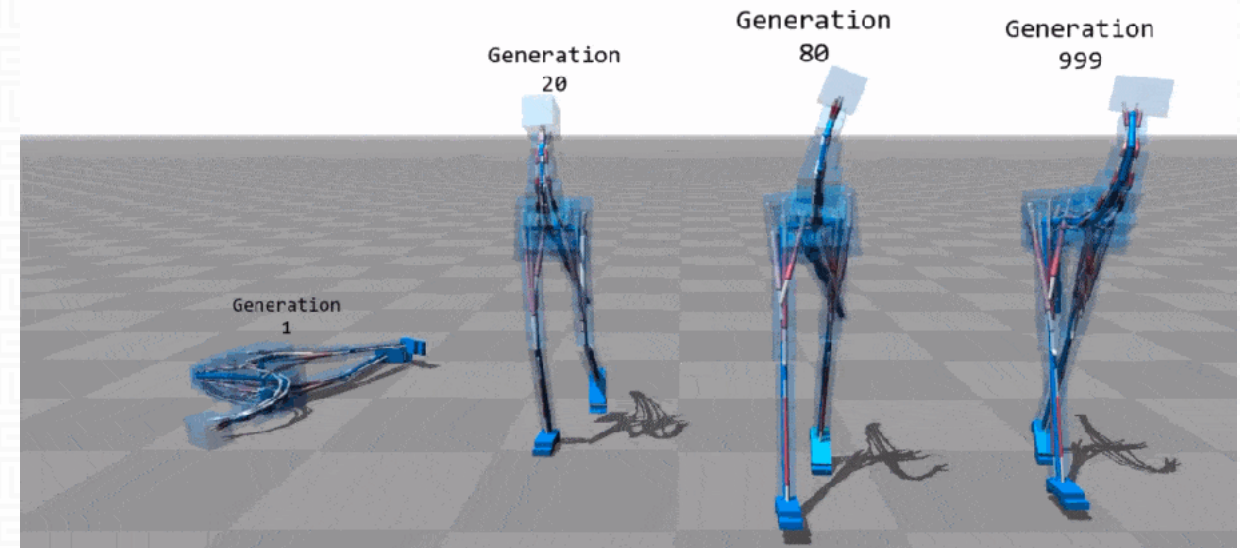


Image from: <https://www.r-bloggers.com/feature-selection-using-genetic-algorithms-in-r/>

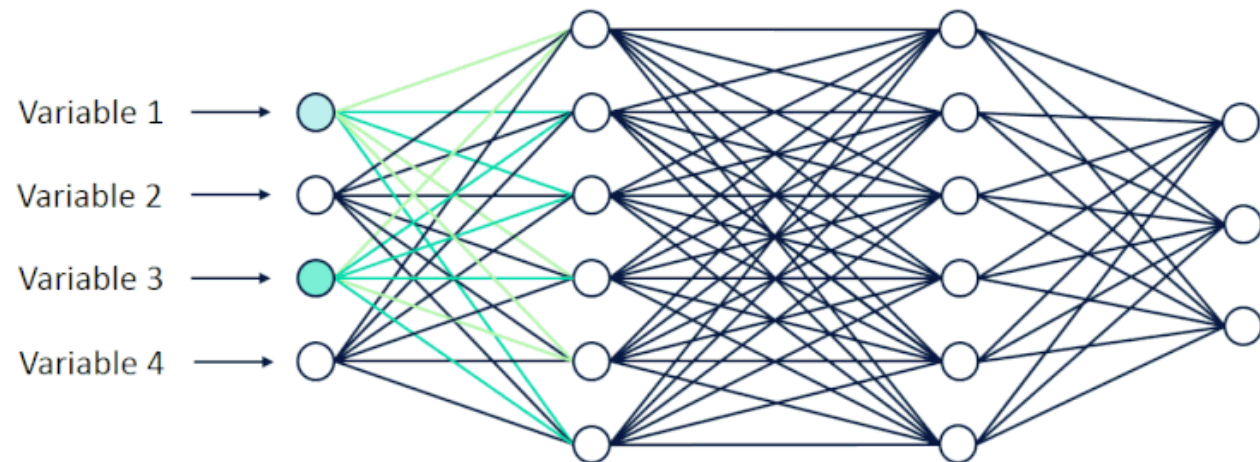
ARTIFICIAL NEURAL NETWORKS (ANN)

Invented early in computing history, yet finally gaining popularity recently thanks to blossom of computation power

Supervised or unsupervised learning?

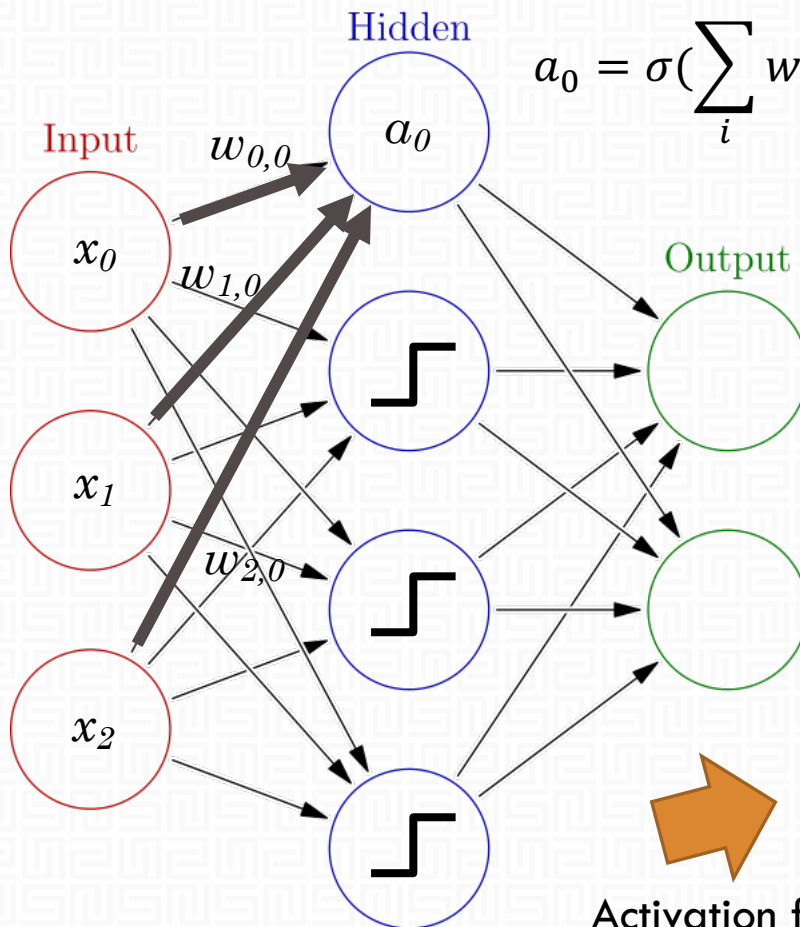
- Trained by examples and cost functions
- Discovering latent structures within unlabelled, unstructured data
- Epochs of iterations

Image from: <https://community.alteryx.com/t5/Data-Science-Blog/It-s-a-No-Brainer-An-Introduction-to-Neural-Networks/ba-p/300479>



A VERY SIMPLE NEURAL NETWORK

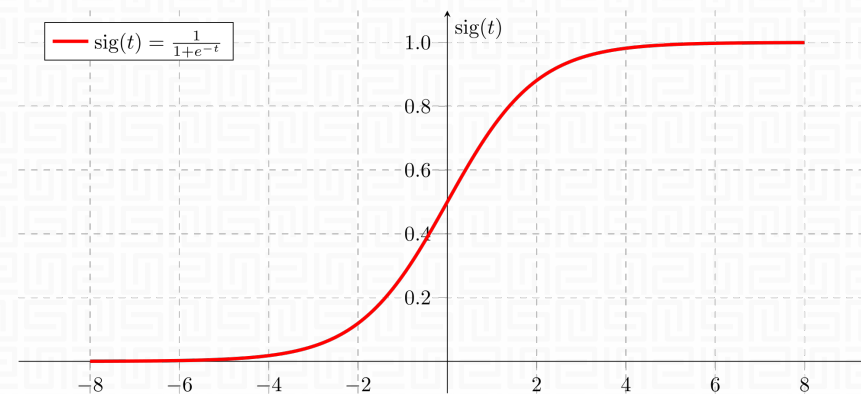
Original image from: https://en.wikipedia.org/wiki/Artificial_neural_network



$$a_0 = \sigma\left(\sum_i w_{i,0}x_i + \text{bias}\right)$$

$$\begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} = \sigma\left(\begin{bmatrix} w_{0,0} & \cdots & w_{0,n} \\ \vdots & \ddots & \vdots \\ w_{k,0} & \cdots & w_{k,n} \end{bmatrix} \begin{bmatrix} x_0 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} b_0 \\ \vdots \\ b_n \end{bmatrix}\right)$$

$$\mathbf{a} = \sigma(\mathbf{w}\mathbf{x} + \mathbf{b})$$



Activation function:
e.g. sigmoid

Interested? Watch more:

<https://youtube.com/watch?v=aircAruvnKk>

A mostly complete chart of Neural Networks

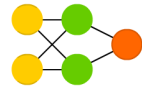
©2019 Fjodor van Veen & Stefan Leijnen asimovinstitute.org

-  Input Cell
-  Backfed Input Cell
-  Noisy Input Cell
-  Hidden Cell
-  Probabilistic Hidden Cell
-  Spiking Hidden Cell
-  Capsule Cell
-  Output Cell
-  Match Input Output Cell
-  Recurrent Cell
-  Memory Cell
-  Gated Memory Cell
-  Kernel
-  Convolution or Pool

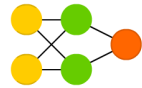
Perceptron (P)



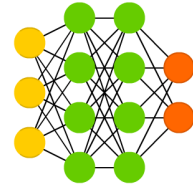
Feed Forward (FF)



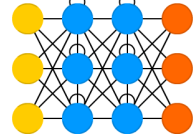
Radial Basis Network (RBF)



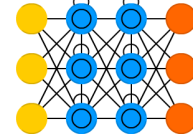
Deep Feed Forward (DFF)



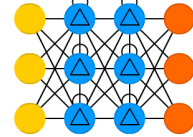
Recurrent Neural Network (RNN)



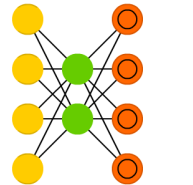
Long / Short Term Memory (LSTM)



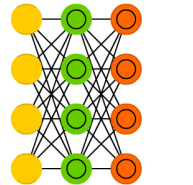
Gated Recurrent Unit (GRU)



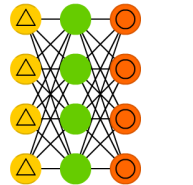
Auto Encoder (AE)



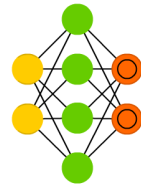
Variational AE (VAE)



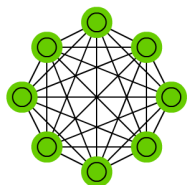
Denoising AE (DAE)



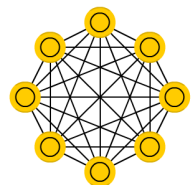
Sparse AE (SAE)



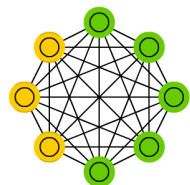
Markov Chain (MC)



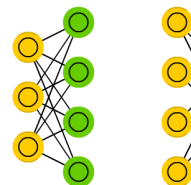
Hopfield Network (HN)



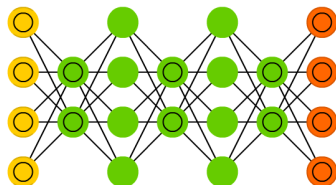
Boltzmann Machine (BM)



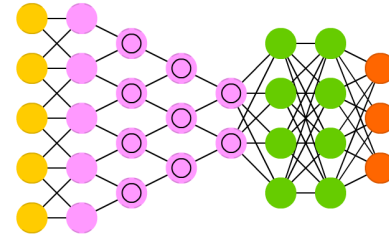
Restricted BM (RBM)



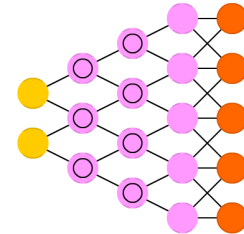
Deep Belief Network (DBN)



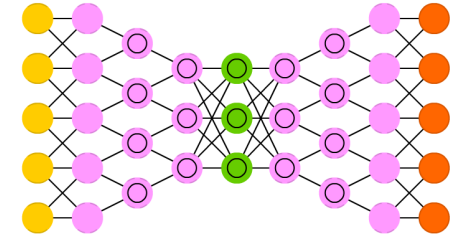
Deep Convolutional Network (DCN)



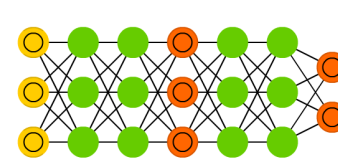
Deconvolutional Network (DN)



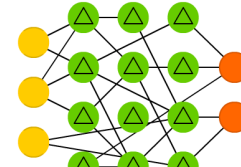
Deep Convolutional Inverse Graphics Network (DCIGN)



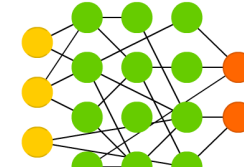
Generative Adversarial Network (GAN)



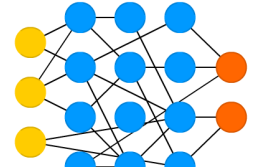
Liquid State Machine (LSM)



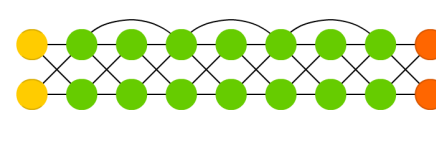
Extreme Learning Machine (ELM)



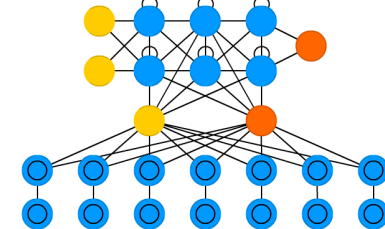
Echo State Network (ESN)



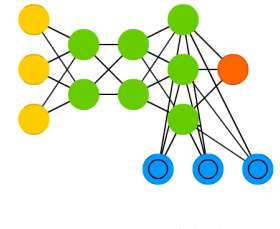
Deep Residual Network (DRN)



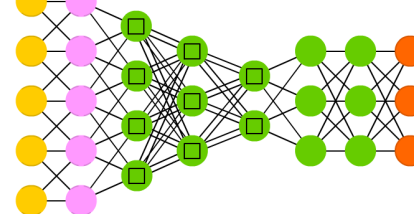
Differentiable Neural Computer (DNC)



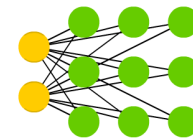
Neural Turing Machine (NTM)



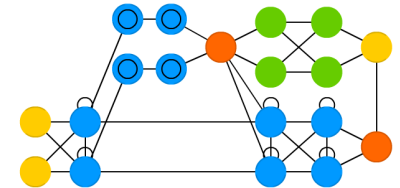
Capsule Network (CN)



Kohonen Network (KN)



Attention Network (AN)



Just too many of them...

RECURRENT NEURAL NETWORK (RNN)

In neural networks, neurons obtain input from previous layer

- With a weighting to decide how much to receive
- Weights are adjusted by training

RNN: in addition to previous layers, gather information also from the previous state of itself

- Suitable for data with time information

Long/Short-term Memory (LSTM)

- Specific kind of RNN with “forget” rate to decide importance of history

Watch this video for more:

- <https://youtube.com/watch?v=UNmqTiOnRfg>

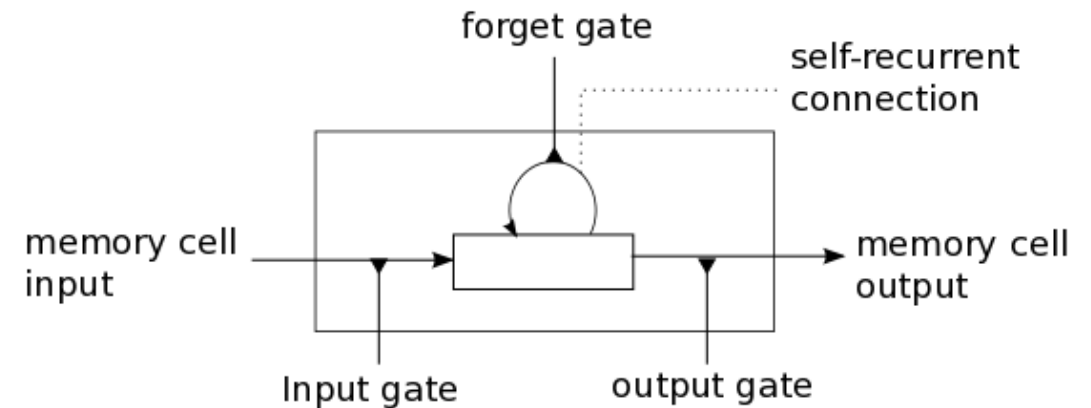


Image from: <http://deeplearning.net/tutorial/lstm.html>

DEEP LEARNING

Deep: *multiple layers between input and output layers*

- Layers of abstraction: hardly understandable by human

Too arbitrary? Heuristic?

- Optimality
- Completeness
- Accuracy and precision

Deep learning frameworks

- TensorFlow
- PyTorch
- MATLAB Deep Learning Toolbox
- See:
 - https://en.wikipedia.org/wiki/Comparison_of_deep_learning_software

WHAT ARE ALL THESE TO DO WITH MUSIC?

Recommendation

- <https://towardsdatascience.com/music-to-my-ears-an-unsupervised-approach-to-user-specific-song-recommendation-6c291acc2c12>

Classification

- <https://medium.com/@juliendespois/finding-the-genre-of-a-song-with-deep-learning-da8f59a61194>
- https://medium.com/@navdeepsingh_2336/identifying-the-genre-of-a-song-with-neural-networks-851db89c42f0

Composition!

- <https://towardsdatascience.com/making-music-with-machine-learning-908ff1b57636>

LECTURE REVIEW

We have discussed:

- Basic ideas of Machine Learning
- Supervised vs. unsupervised learning
- Reinforcement learning
- Artificial neural networks
 - RNN/LSTM

READ FURTHER

Useful websites:

- https://vas3k.com/blog/machine_learning
- <https://medium.com/machine-learning-for-humans>
- https://en.wikipedia.org/wiki/Machine_learning