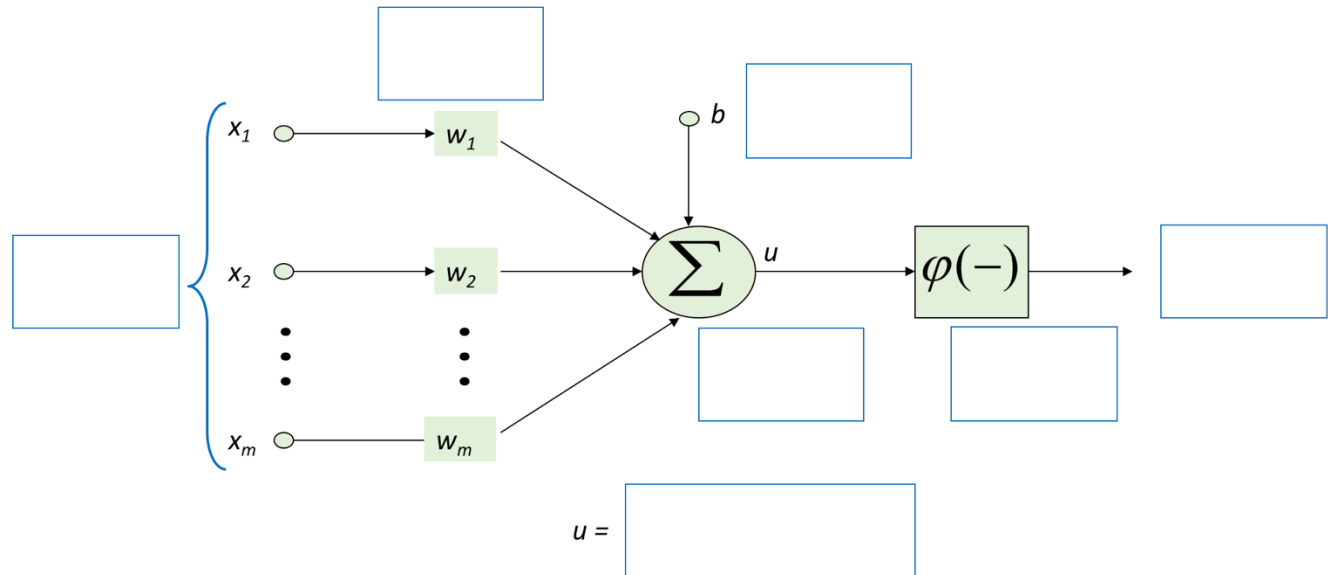


**CSE 40647/60647 Data Science (Spring 2018)**  
**Lecture 12: Advanced Classification: Neural Networks**

**1. Artificial Neurons**



**2. Activation function: Non-linear, parameterized function**

**a. Logistic functions**

$f(x) =$

**Shape:**

if  $x \in (-\infty, +\infty)$ , then  $f(x) \in (\_, \_)$

**b. Sigmoid functions**

$f(x) =$

**Shape:**

if  $x \in (-\infty, +\infty)$ , then  $f(x) \in (\_, \_)$

**c. Hyperbolic tangent function**

$f(x) =$

**Shape:**

if  $x \in (-\infty, +\infty)$ , then  $f(x) \in (\_, \_)$

if  $x \in [0, +\infty)$ , then  $f(x) \in (\_, \_)$

### 3. Example: Classifying a paper as a KDD paper or a non-KDD paper.

	"data mining"	"web search"	"click through rate"	"pattern"	Label (KDD?)
PID <sub>1</sub>	1	0	0	1	1
PID <sub>2</sub>	0	1	1	0	0
PID <sub>3</sub>	1	1	0	1	1

**Optimization function:**

**Partial derivatives:**

**Initialized weights:** [ $w_0=0.1$ ,  $w_{\text{data mining}}=0.1$ ,  $0.1$ ,  $0.1$ ,  $w_{\text{pattern}}=0.1$ ]

**Predicted values on the training instances:**

PID <sub>1</sub>	
PID <sub>2</sub>	
PID <sub>3</sub>	

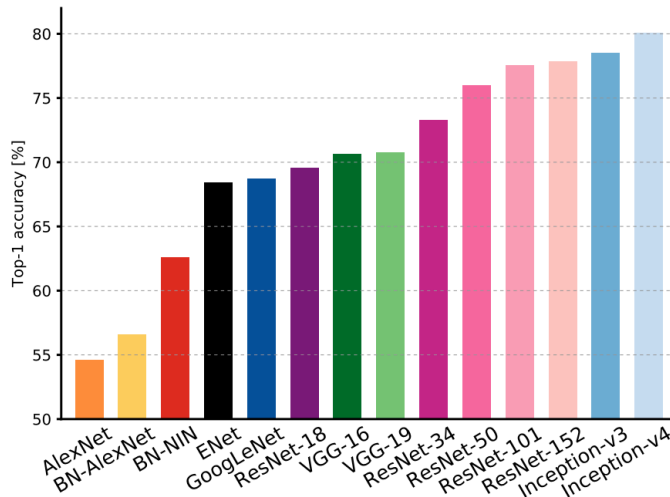
**Gradients:**

**New weights:** [ ]

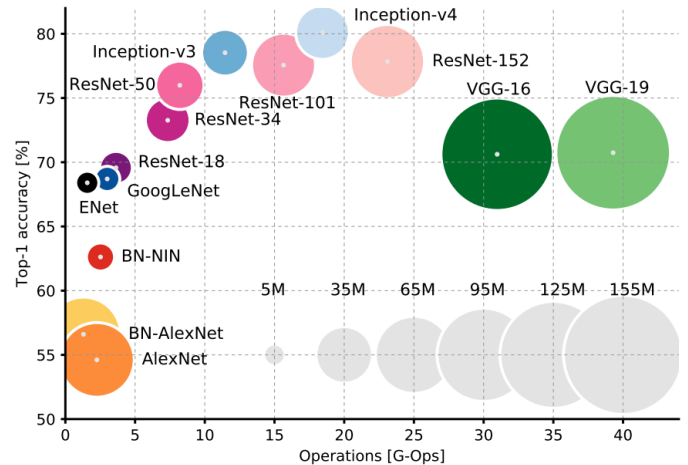
**New predicted values on the training instances:**

PID <sub>1</sub>	
PID <sub>2</sub>	
PID <sub>3</sub>	

# Research: AN ANALYSIS OF DEEP NEURAL NETWORK MODELS FOR PRACTICAL APPLICATIONS



**Figure 1: Top1 vs. network.** Single-crop top-1 validation accuracies for top scoring single-model architectures. We introduce with this chart our choice of colour scheme, which will be used throughout this publication to distinguish effectively different architectures and their correspondent authors. Notice that networks of the same group share the same hue, for example ResNet are all variations of pink.



**Figure 2: Top1 vs. operations, size  $\propto$  parameters.** Top-1 one-crop accuracy versus amount of operations required for a single forward pass. The size of the blobs is proportional to the number of network parameters; a legend is reported in the bottom right corner, spanning from  $5 \times 10^6$  to  $155 \times 10^6$  params. Both these figures share the same y-axis, and the grey dots highlight the centre of the blobs.

A mostly complete chart of

## Neural Networks

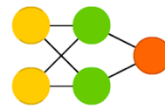
©2016 Fjodor van Veen - asimovinstitute.org

- Backfed Input Cell
- Input Cell
- Noisy Input Cell
- Hidden Cell
- Probablistic Hidden Cell
- Spiking Hidden Cell
- Output Cell
- Match Input Output Cell
- Recurrent Cell
- Memory Cell
- Different Memory Cell
- Kernel
- Convolution or Pool

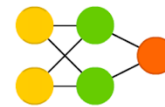
Perceptron (P)



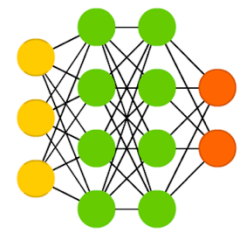
Feed Forward (FF)



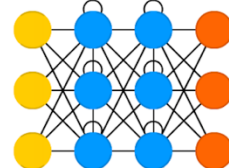
Radial Basis Network (RBF)



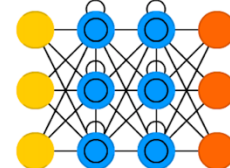
Deep Feed Forward (DFF)



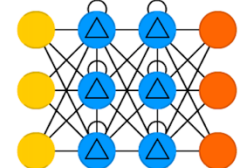
Recurrent Neural Network (RNN)



Long / Short Term Memory (LSTM)



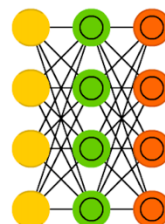
Gated Recurrent Unit (GRU)



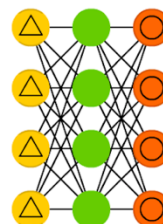
Auto Encoder (AE)



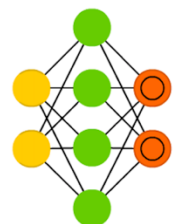
Variational AE (VAE)



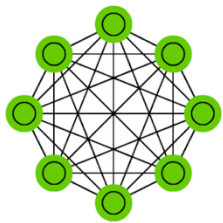
Denoising AE (DAE)



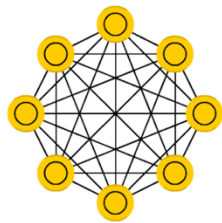
Sparse AE (SAE)



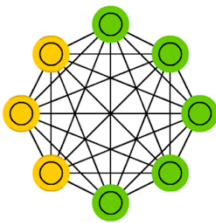
Markov Chain (MC)



Hopfield Network (HN)



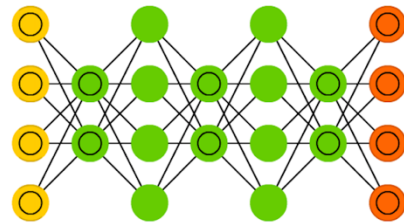
Boltzmann Machine (BM)



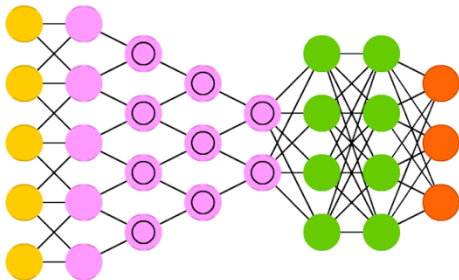
Restricted BM (RBM)



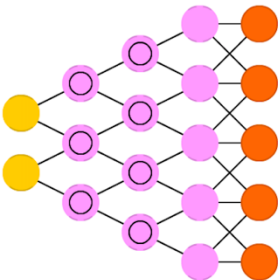
Deep Belief Network (DBN)



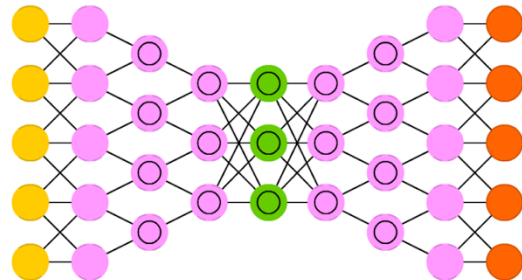
Deep Convolutional Network (DCN)



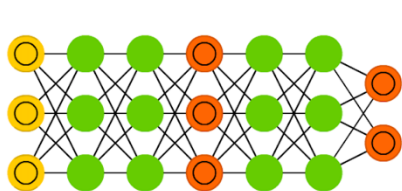
Deconvolutional Network (DN)



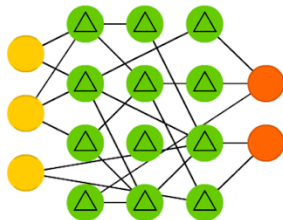
Deep Convolutional Inverse Graphics Network (DCIGN)



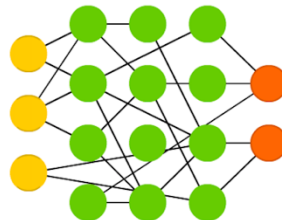
Generative Adversarial Network (GAN)



Liquid State Machine (LSM)



Extreme Learning Machine (ELM)



Echo State Network (ESN)

