

# Uncommon Neural Architectures for Common Data Science Problems

Samiran Roy

Senior Lead Data Scientist at Envestnet | Yodlee, Masters at IIT Bombay

LinkedIn: samiranroy

Slides: <http://bit.ly/NeuralPPT>

# Tentative outline

Philosophy (7 Mins)

Deep Learning Review (10 Mins)

Semantic Hashing (10 Mins)

Siamese Networks (10 Mins)

Multi-Modal Deep Learning (10 Mins)

Q&A (5 Mins)

# Mentimeter

# Data Scientist: Two Perspectives



**Researcher**

Implements research papers

Chases state of the art benchmarks

Seeks fundamental understanding



**Industry Guy**

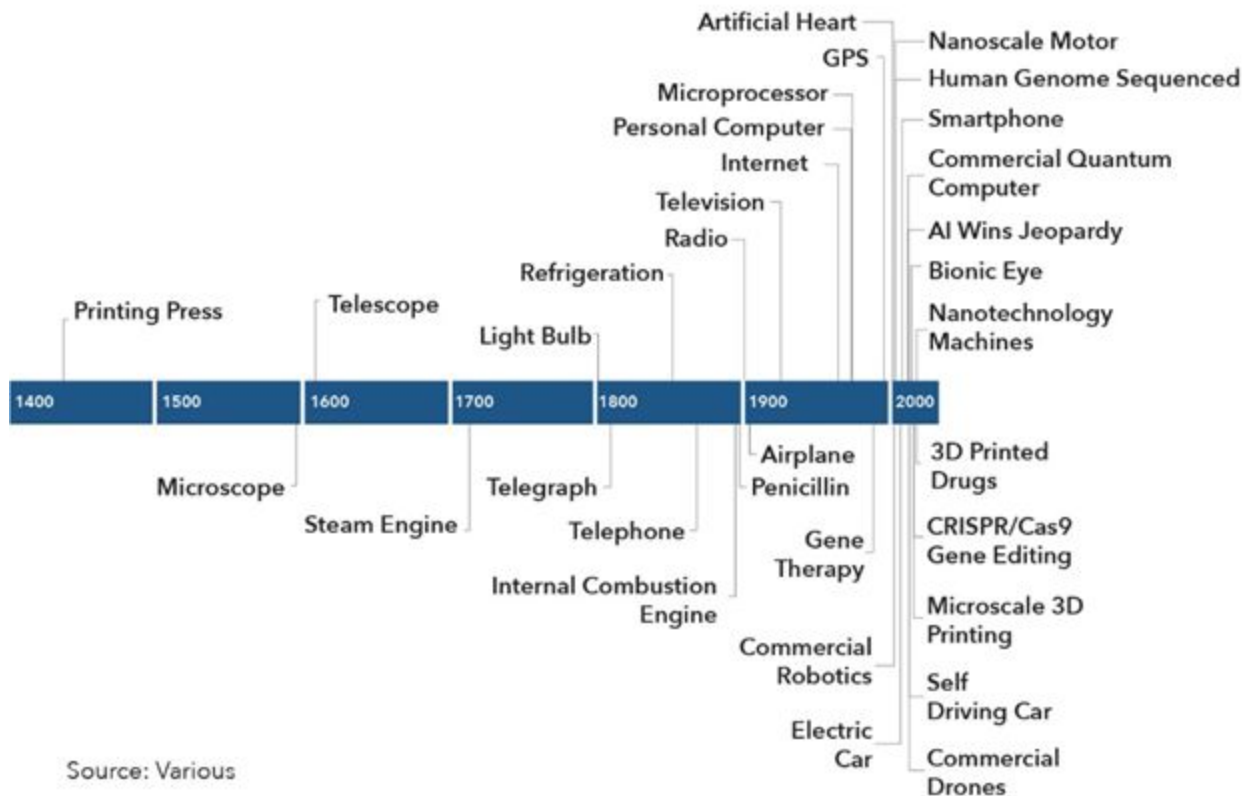
Looks for tools and libraries

Chases buzzwords

Models are a black box

# INCREASING RATE OF CHANGE

FREQUENCY OF KEY TECHNOLOGICAL DEVELOPMENTS

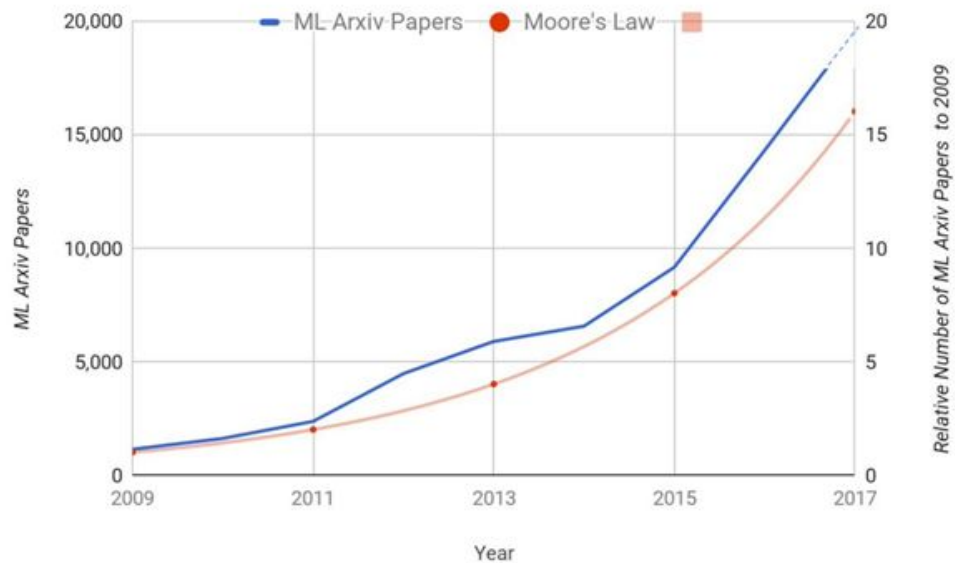


# Job Trends from Indeed.com

— "Data Scientist"



# ML Arxiv Papers per Year



**MACHINE LEARNING**

**MACHINE LEARNING EVERYWHERE**



# General Rules of Thumb

# General Rules of Thumb

We should be great at one thing

# General Rules of Thumb

We should be great at one thing

We should be very good at all the basic things

# General Rules of Thumb

We should be great at one thing

We should be very good at all the basic things

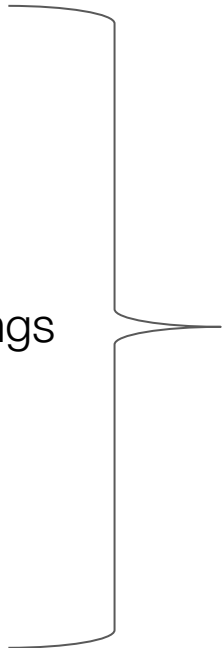
We should keep an eye on everything

# General Rules of Thumb

We should be great at one thing

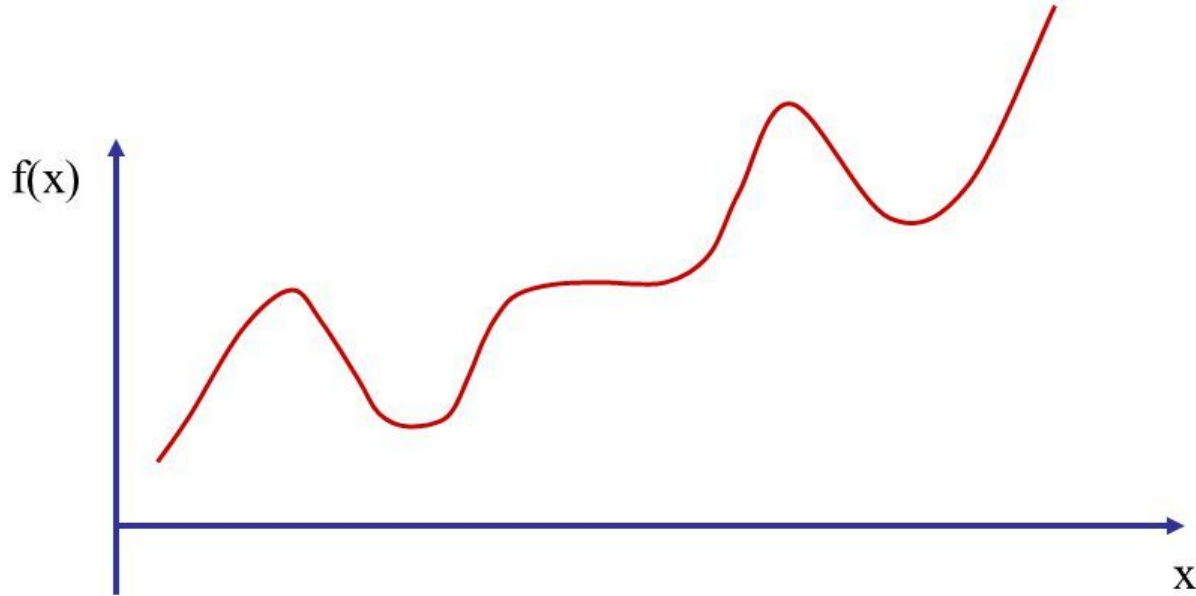
We should be very good at all the basic things

We should keep an eye on everything



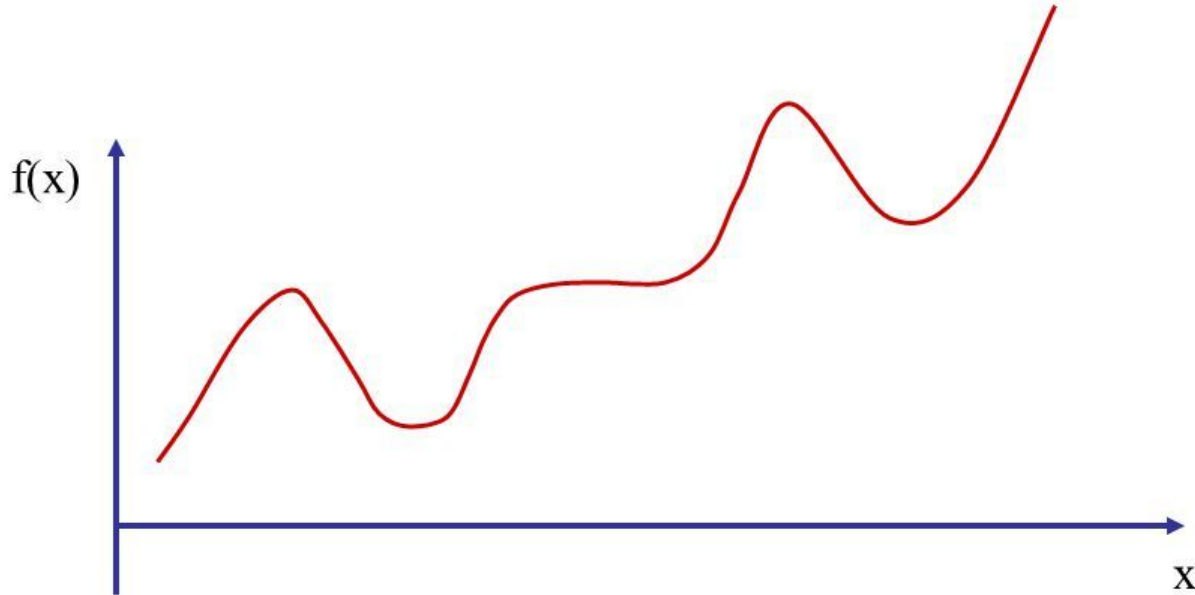
Keeping a strong focus on  
the underlying mathematics

# Deep Learning Review



Can a 1 hidden layer neural network represent any arbitrary function?

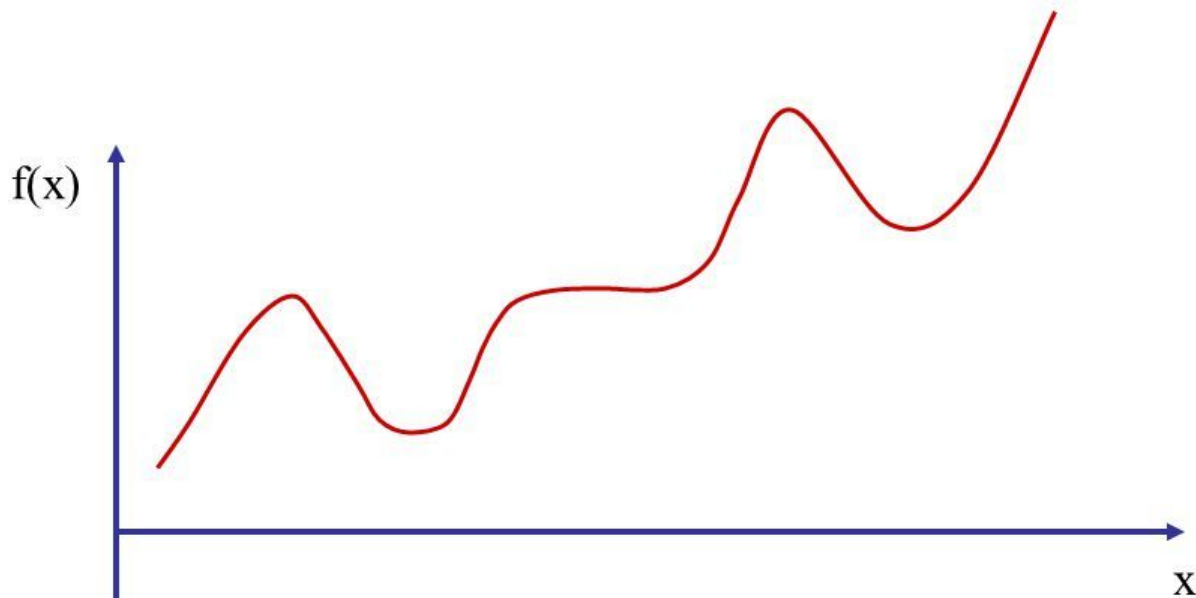
# Deep Learning Review



Can a 1 hidden layer neural network represent any arbitrary function?

Yes! Universal Approximation Theorem

# Learnability

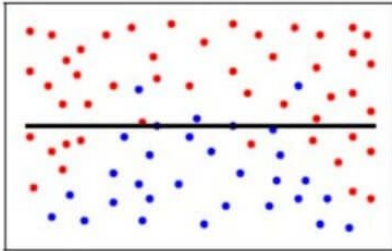


Can a 1 hidden layer neural network *learn* any arbitrary function?

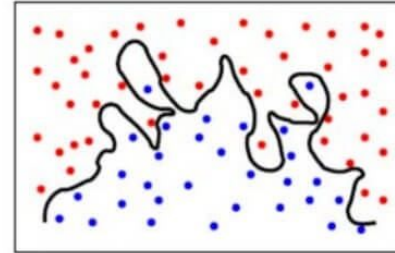


# Overfitting

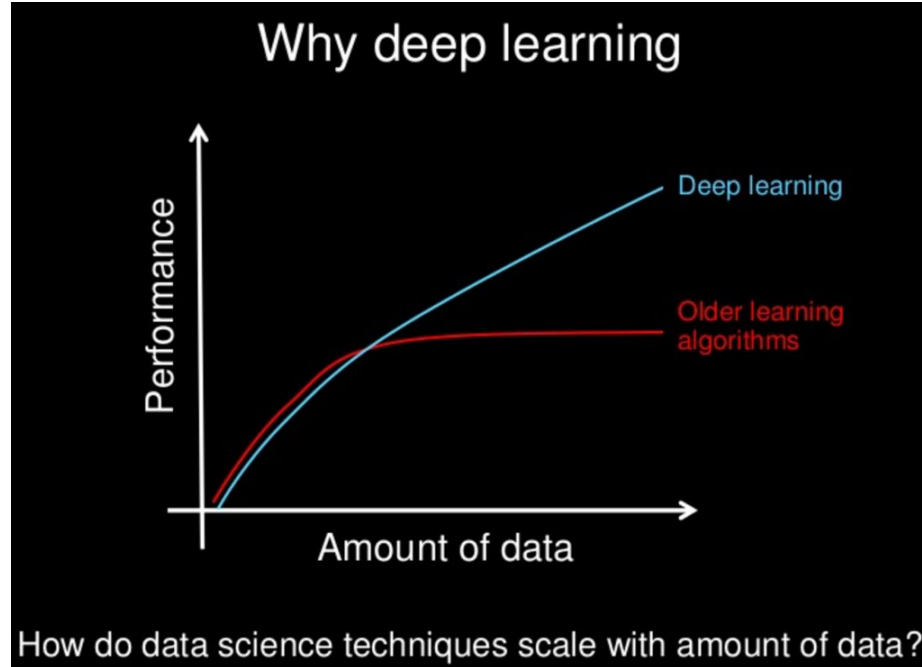
Underfitting



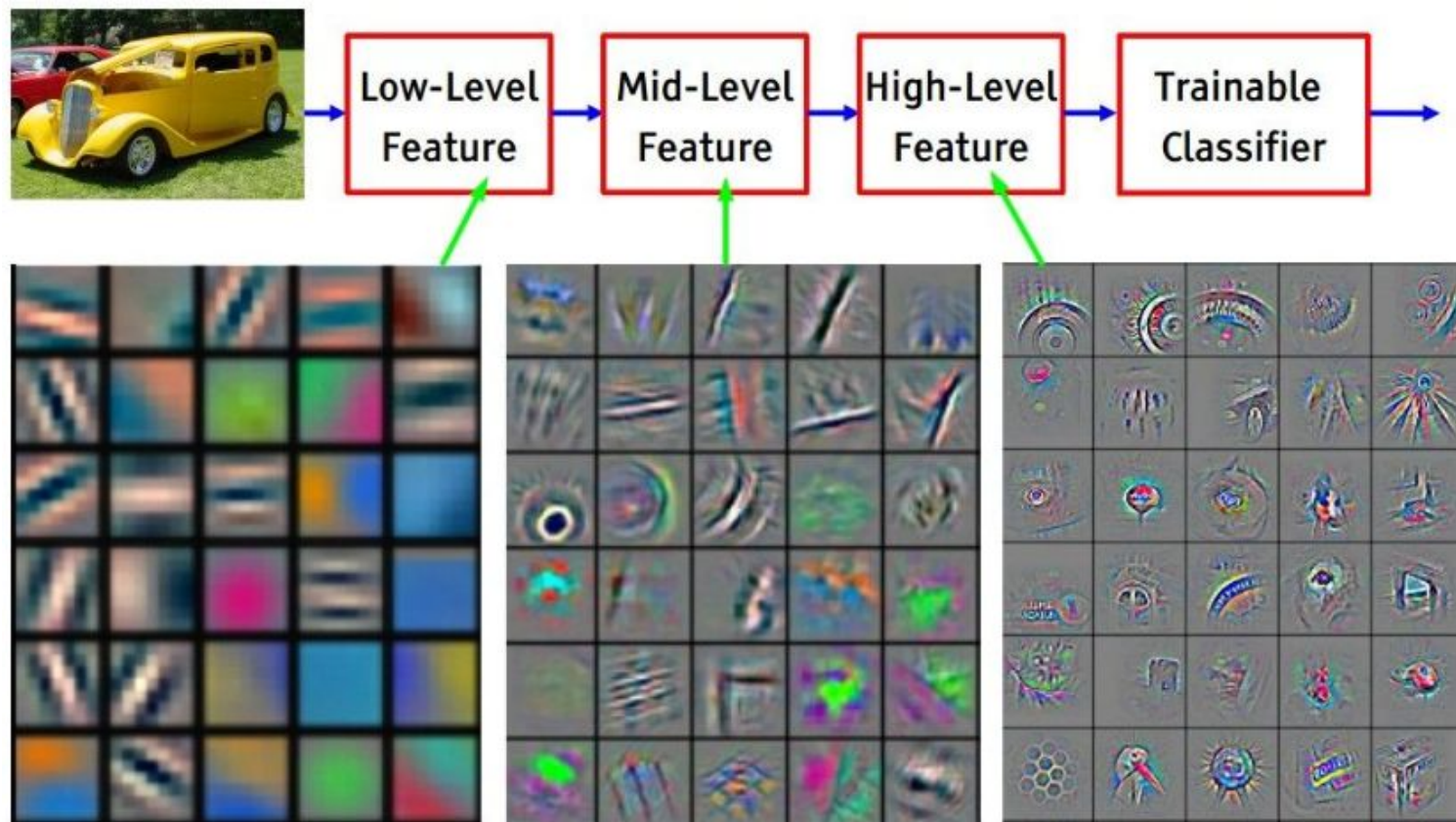
Overfitting



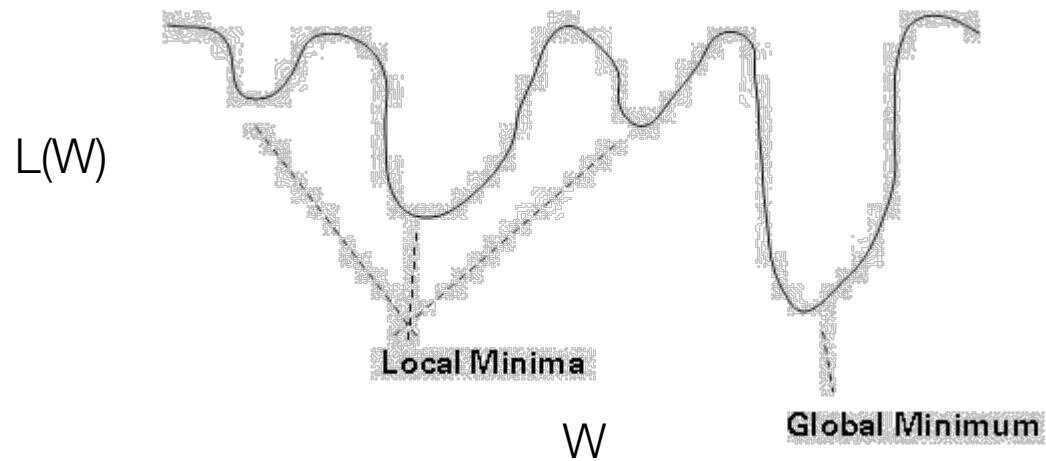
# Deep Learning Review



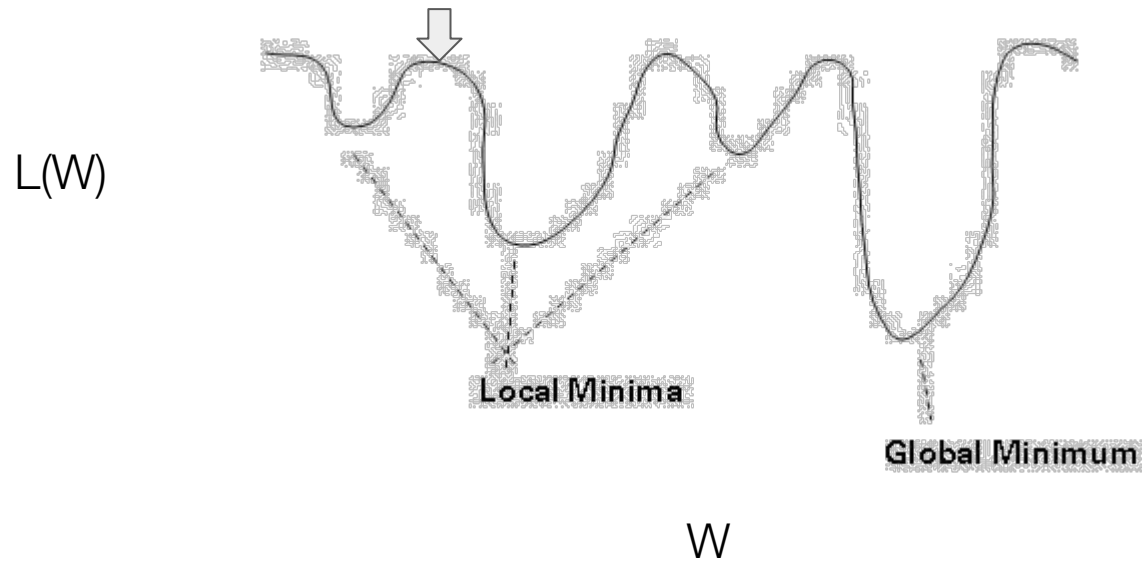
# Deep Learning Review



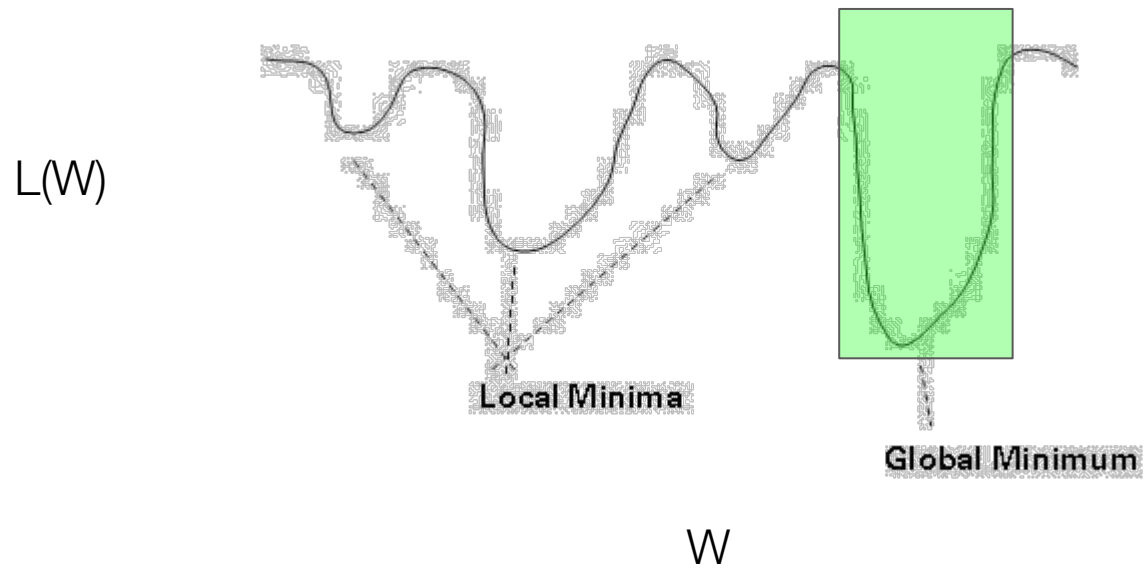
# Initialization



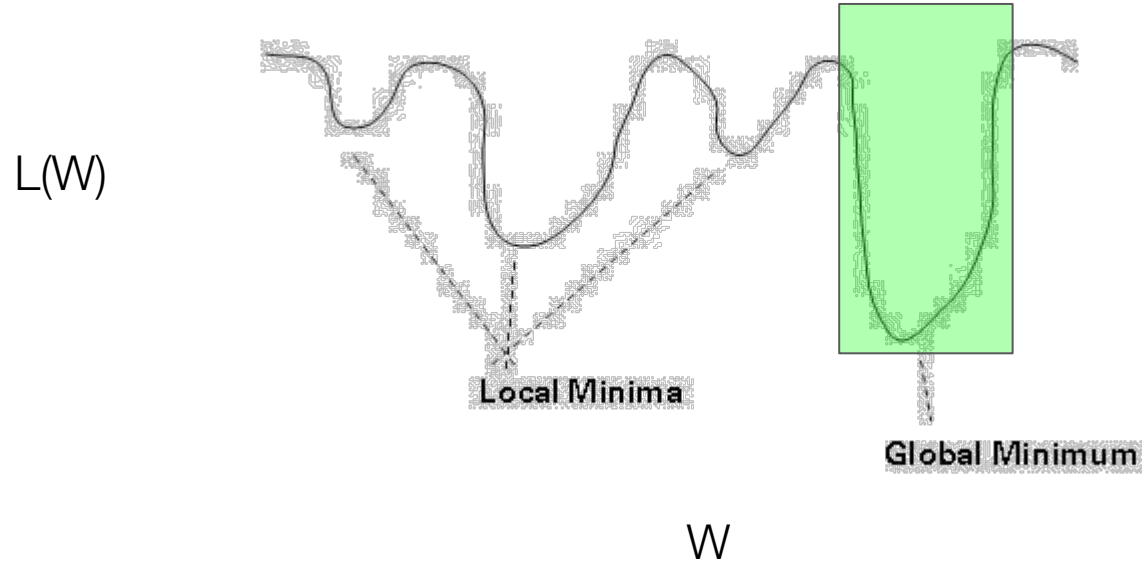
# Initialization



# Initialization

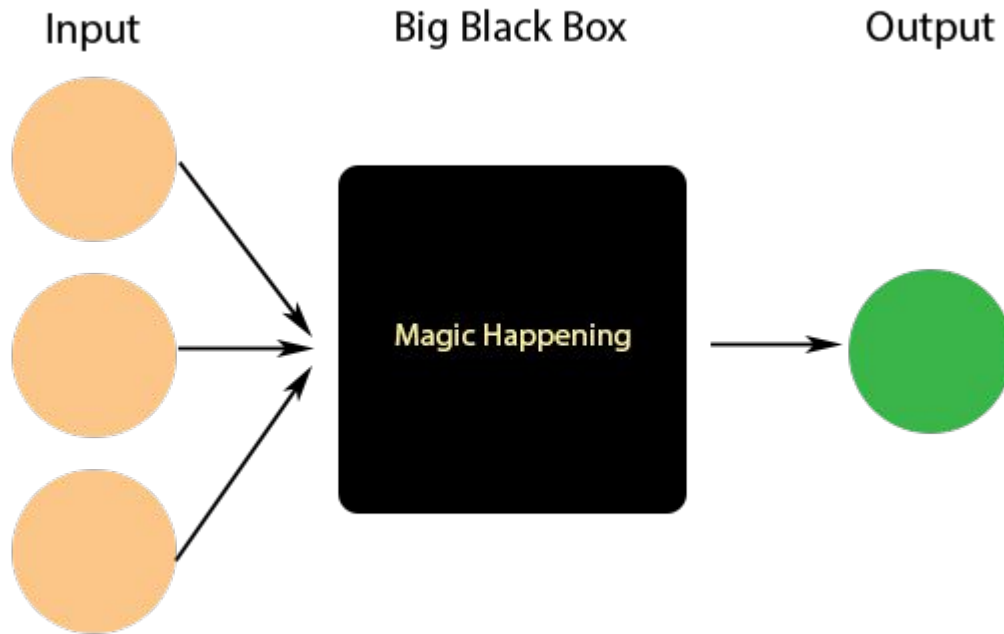


# Initialization



How do you escape local minima in higher dimensional non-convex spaces?

# Uncommon Neural Architectures for Common Data Science Problems





# Architecture Design

Setting:

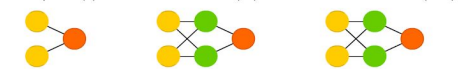
Supervised, Unsupervised, Semi-Supervised, Weakly Supervised, Self-Supervised, Reinforcement Learning, Active Learning, Transfer Learning, Meta Learning, etc..

# Architecture Design

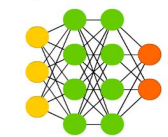
## A mostly complete chart of Neural Networks

©2019 Fjodor van Veen & Stefan Leijnen asimovinstitute.org

Perceptron (P) Feed Forward (FF) Radial Basis Network (RBF)



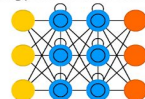
Deep Feed Forward (DFF)



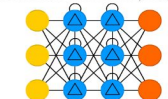
Recurrent Neural Network (RNN)



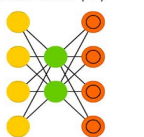
Long / Short Term Memory (LSTM)



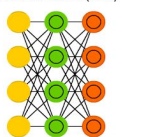
Gated Recurrent Unit (GRU)



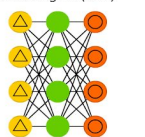
Auto Encoder (AE)



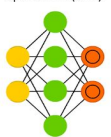
Variational AE (VAE)



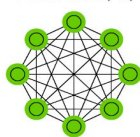
Denosing AE (DAE)



Sparse AE (SAE)



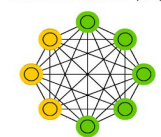
Markov Chain (MC)



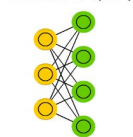
Hopfield Network (HN)



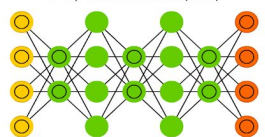
Boltzmann Machine (BM)



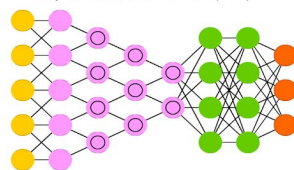
Restricted BM (RBM)



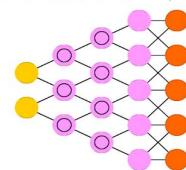
Deep Belief Network (DBN)



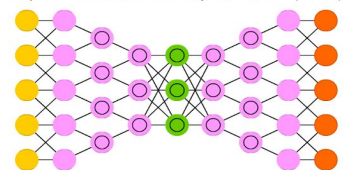
Deep Convolutional Network (DCN)



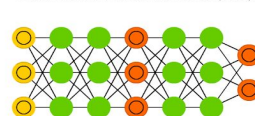
Deconvolutional Network (DN)



Deep Convolutional Inverse Graphics Network (DCIGN)



Generative Adversarial Network (GAN)



Liquid State Machine (LSM)



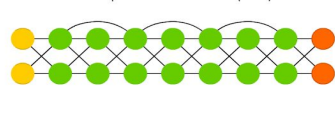
Extreme Learning Machine (ELM)



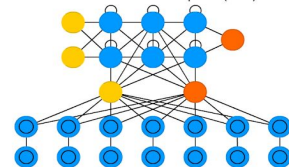
Echo State Network (ESN)



Deep Residual Network (DRN)



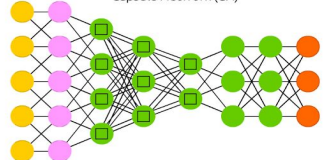
Differentiable Neural Computer (DNC)



Neural Turing Machine (NTM)



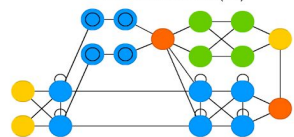
Capsule Network (CN)



Kohonen Network (KN)



Attention Network (AN)



- 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

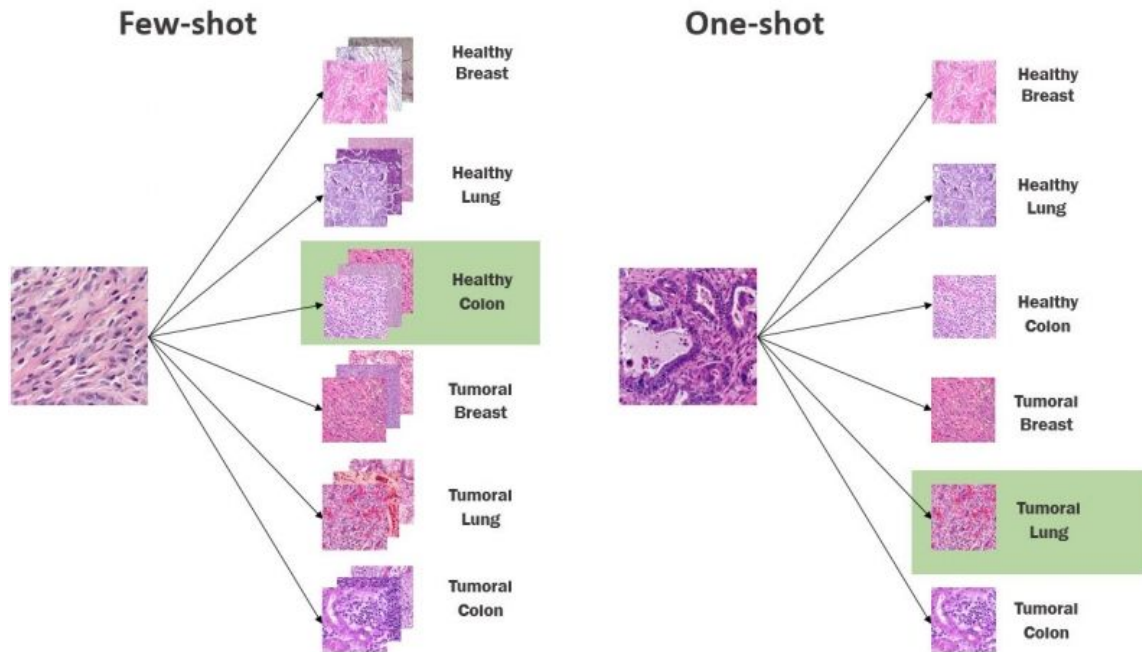
# Problem Statement: Unbound/ Substantial Number of Classes



# Problem Statement: One Shot Learning



# Problem Statement: One Shot Learning



# Siamese Neural Networks - Resources

<https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf>

[https://www.cs.toronto.edu/~ranzato/publications/taigman\\_cvpr14.pdf](https://www.cs.toronto.edu/~ranzato/publications/taigman_cvpr14.pdf)

<https://medium.com/mlreview/implementing-malstm-on-kaggles-quora-question-pairs-competition-8b31b0b16a07>

<https://towardsdatascience.com/image-similarity-using-triplet-loss-3744c0f67973>

# Siamese Networks for Intent Detection

ACH debit mcdonalds purchase samiran roy 20/08 Sand Lake Rd, Orlando, FL

ACH debit MCDNLDS purchase samiran roy 20/08 Sand Lake Rd, Orlando, FL

ACH debit Central Florida Pkwy mcdonalds

ACH credit mcdonalds purchase samiran roy 20/08 Sand Lake Rd, Orlando, FL

ACH debit H&M purchase samiran roy 20/08 Sand Lake Rd, Orlando, FL

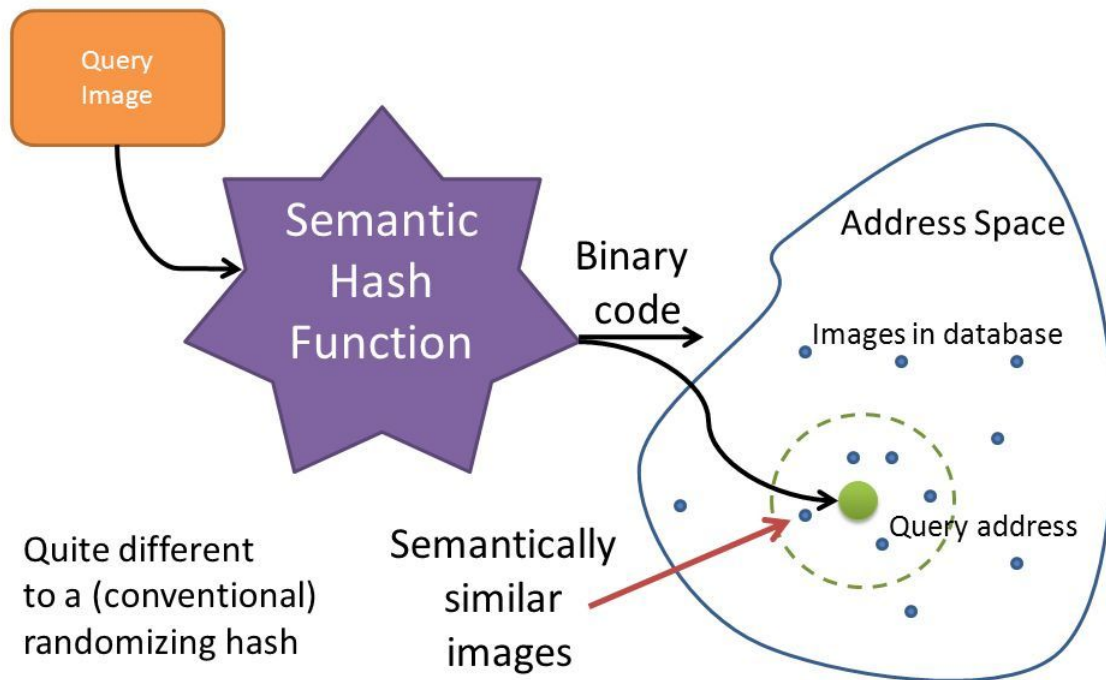
Refund Mcdonalds samiran roy 20/08 Sand Lake Rd, Orlando, FL

# Approximate Similarity Search





# Semantic Hashing



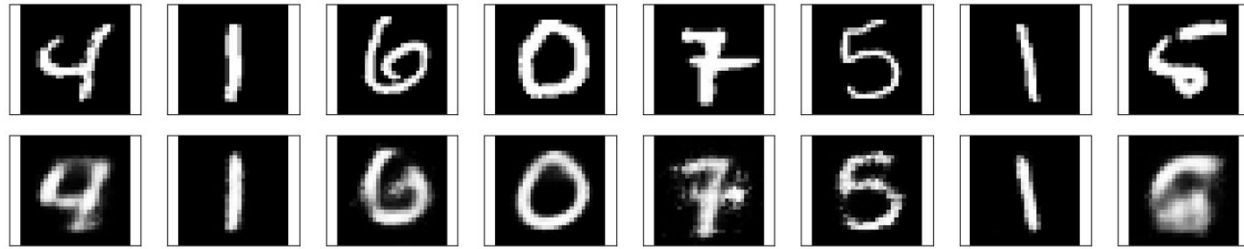
# Semantic Hashing

Dataset	LabelMe	Web
# datapoints	$2 \times 10^4$	$1.29 \times 10^7$
Gist vector dim.	512	384

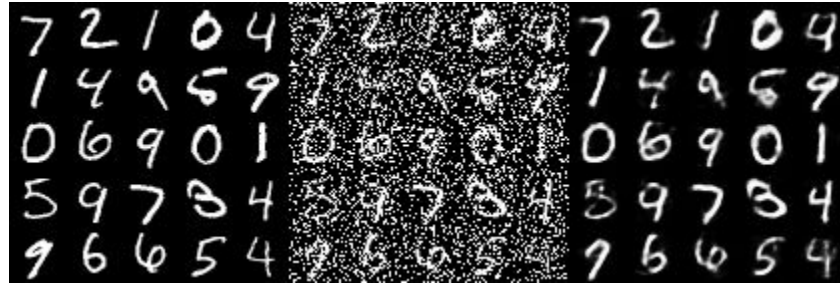
  

Method	Time (s)	Time (s)
Spill tree - Gist vector	1.05	-
Brute force - Gist vector	0.38	-
Brute force - 30 bit binary	$4.3 \times 10^{-4}$	0.146
" - 30 bit binary, M/T	$2.7 \times 10^{-4}$	0.074
Brute force - 256 bit binary	$1.4 \times 10^{-3}$	0.75
" - 256 bit binary, M/T	$4.7 \times 10^{-4}$	0.23
Sem. Hashing - 30 bit binary	$6 \times 10^{-6}$	$6 \times 10^{-6}$

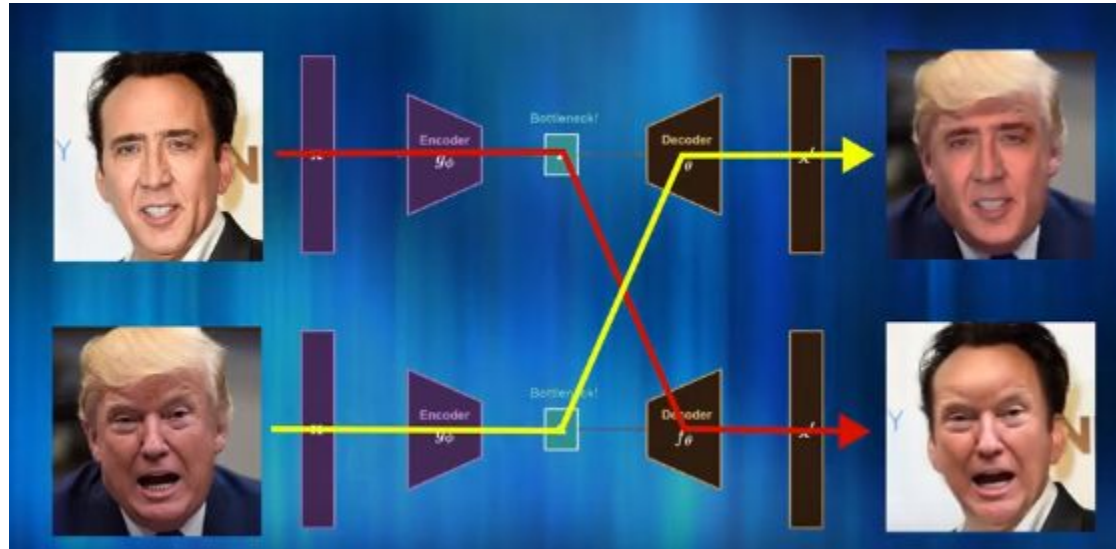
# Autoencoders - Reconstructed Images



# Autoencoders



# Autoencoders



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

# Semantic Hashing - Resources

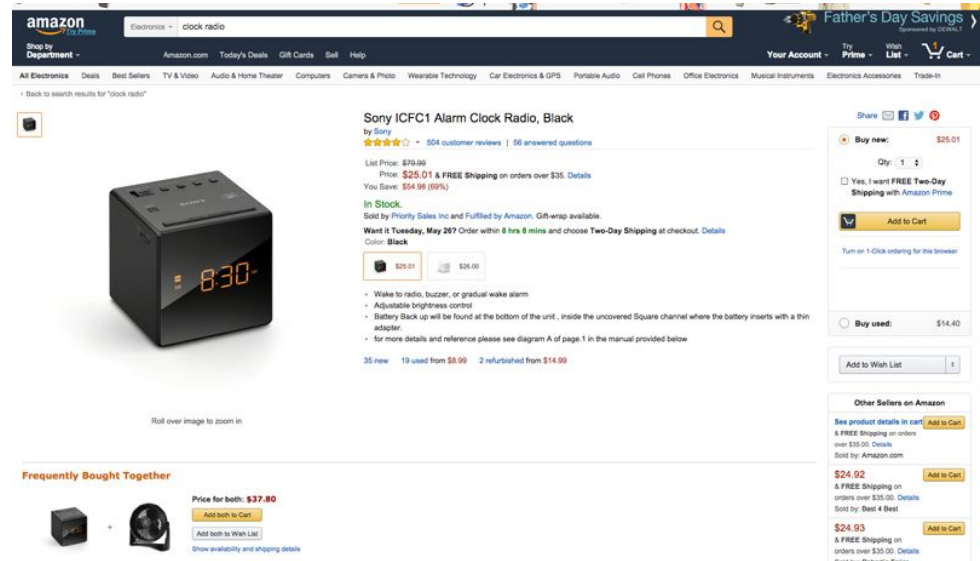
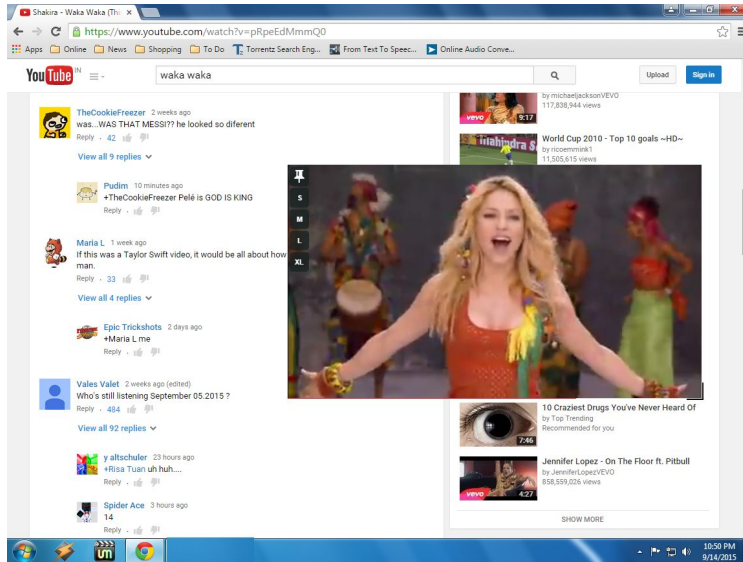
[https://www.cs.utoronto.ca/~rsalakhu/papers/semantic\\_final.pdf](https://www.cs.utoronto.ca/~rsalakhu/papers/semantic_final.pdf)

<https://www.youtube.com/watch?v=uaaqyVS9-rM&t=442s>

<https://arxiv.org/pdf/1708.03436.pdf>

<https://github.com/erikbern/ann-benchmarks>

# Multi-Modal Deep Learning





# Flickr



dog

dogs

Chocolate Lab

Labrador Retriever

Frisbee

game

lazy

energy

chair

man

Dana Point

summer

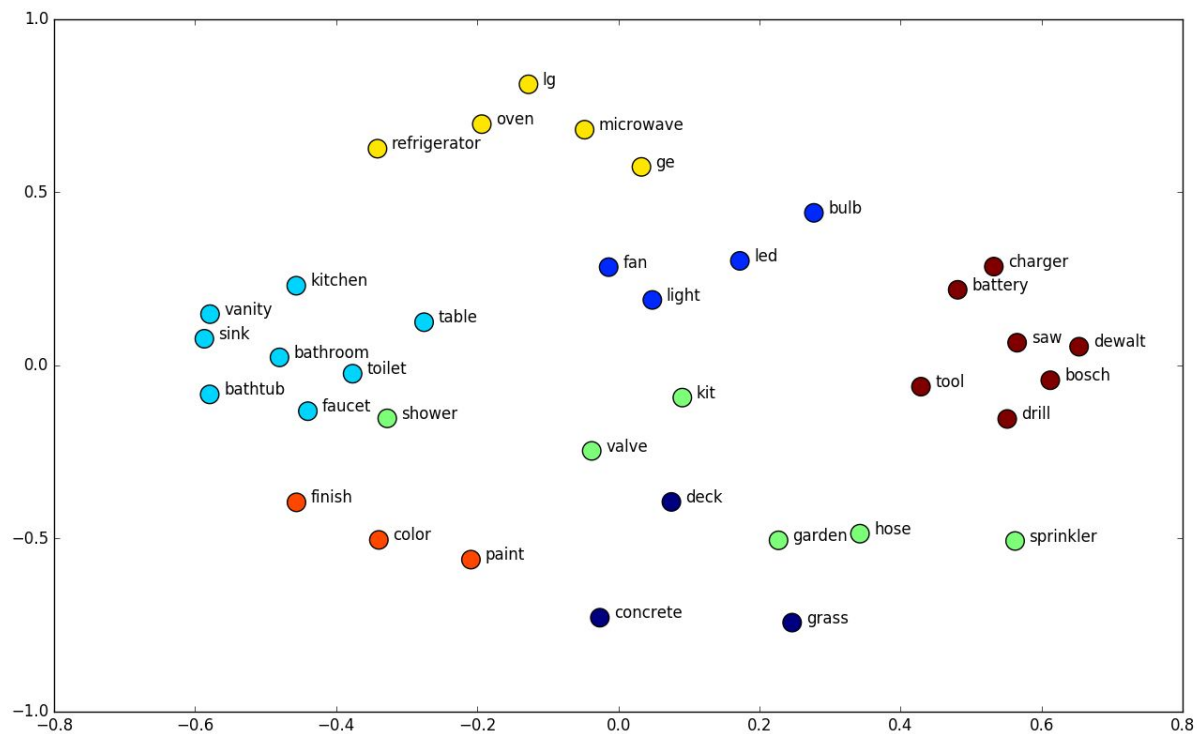
hot

games

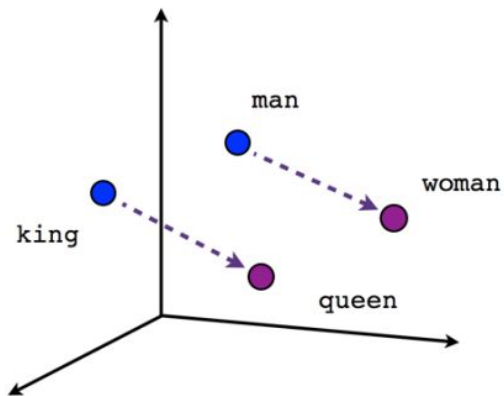
exercising



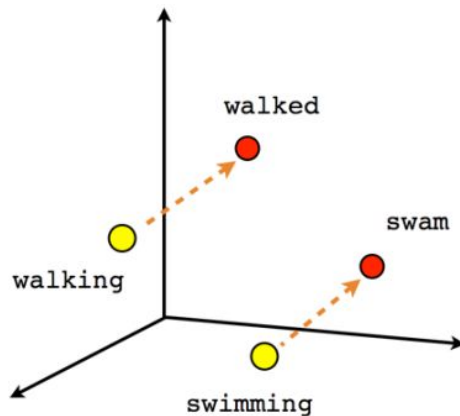
# Word Embeddings



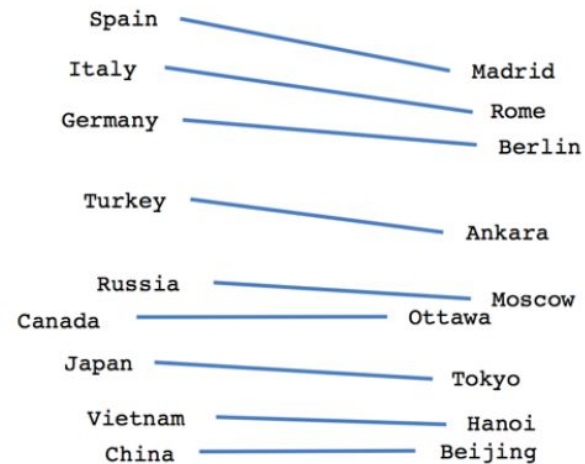
# Vector Arithmetic



Male-Female



Verb tense



Country-Capital

# Number of Parameters

Layer (type)	Output Shape	Param #	Connected to
convolution2d_11 (Convolution2D)	(None, 32, 76L, 76L)	832	convolution2d_input_6[0][0]
activation_21 (Activation)	(None, 32, 76L, 76L)	0	convolution2d_11[0][0]
maxpooling2d_11 (MaxPooling2D)	(None, 32, 38L, 38L)	0	activation_21[0][0]
convolution2d_12 (Convolution2D)	(None, 64, 34L, 34L)	51264	maxpooling2d_11[0][0]
activation_22 (Activation)	(None, 64, 34L, 34L)	0	convolution2d_12[0][0]
maxpooling2d_12 (MaxPooling2D)	(None, 64, 17L, 17L)	0	activation_22[0][0]
dropout_11 (Dropout)	(None, 64, 17L, 17L)	0	maxpooling2d_12[0][0]
flatten_6 (Flatten)	(None, 18496)	0	dropout_11[0][0]
dense_11 (Dense)	(None, 100)	1849700	flatten_6[0][0]
activation_23 (Activation)	(None, 100)	0	dense_11[0][0]
dropout_12 (Dropout)	(None, 100)	0	activation_23[0][0]
dense_12 (Dense)	(None, 2)	202	dropout_12[0][0]
activation_24 (Activation)	(None, 2)	0	dense_12[0][0]
Total params: 1901998			

# Results

## Flickr tagging: generate tags given images

Given	Generated	Given	Generated
	dog, cat, pet, kitten, puppy, ginger, tongue, kitty, dogs, furry		insect, butterfly, insects, bug, butterflies, lepidoptera
	sea, france, boat, mer, beach, river, bretagne, plage, brittany		graffiti, streetart, stencil, sticker, urbanart, graff, sanfrancisco
	portrait, child, kid, ritratto, kids, children, boy, cute, boys, italy		canada, nature, sunrise, ontario, fog, mist, bc, morning

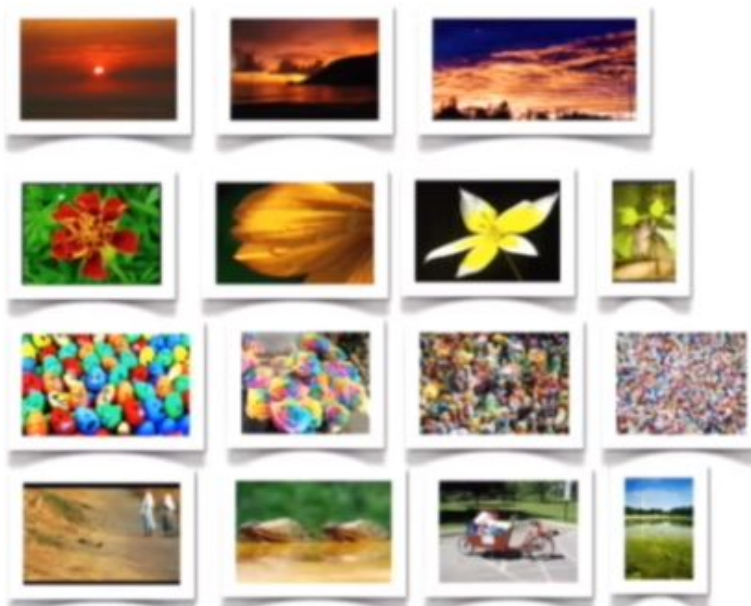
# Results

## Flickr tagging: find images given tags

Given

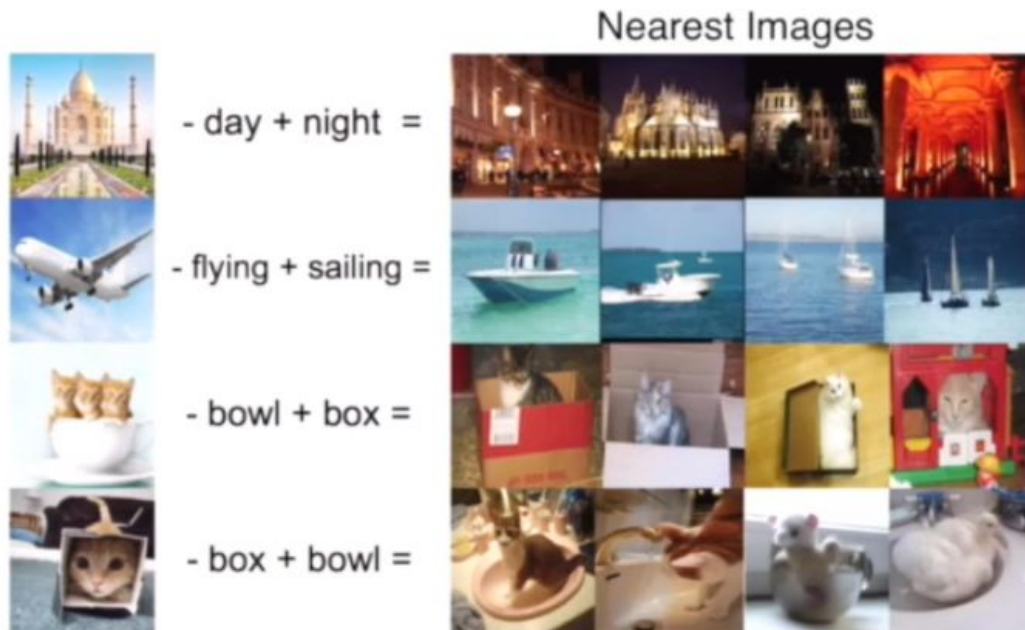
water, red,  
sunset

Retrieved



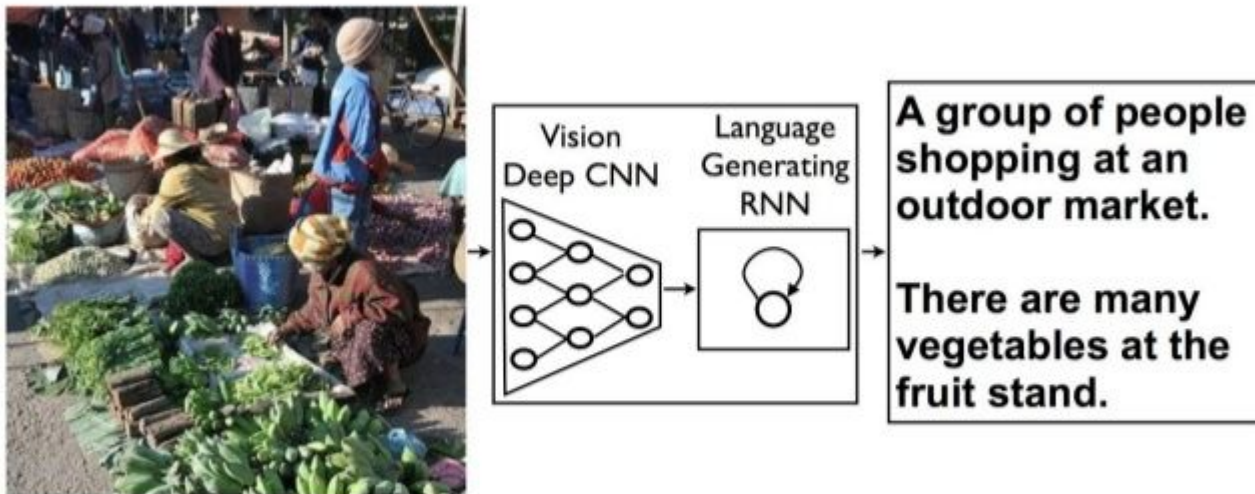
# Results

## Flickr tagging: multimodal arithmetic



# Results

## Caption Generation



<http://arxiv.org/abs/1411.4555> "Show and Tell: A Neural Image Caption Generator"



# Results



\*We were barely able to catch the breeze at the beach , and it felt as if someone stepped out of my mind . She was in love with him for the first time in months , so she had no intention of escaping . The sun had risen from the ocean , making her feel more alive than normal . She 's beautiful , but the truth is that I do n't know what to do . The sun was just starting to fade away , leaving people scattered around the Atlantic Ocean . I d seen the men in his life , who guided me at the beach once more .\*

Source: <https://github.com/ryankiros/neural-storyteller>



# Multi Modal Deep Learning - Resources

<https://www.youtube.com/watch?v=6QewMQT4iMM>

<https://arxiv.org/abs/1505.00487>

<https://arxiv.org/abs/1605.05396>

<http://soundnet.csail.mit.edu/>

# Thanks!

<https://www.youtube.com/watch?v=cQ54GDm1eL0>