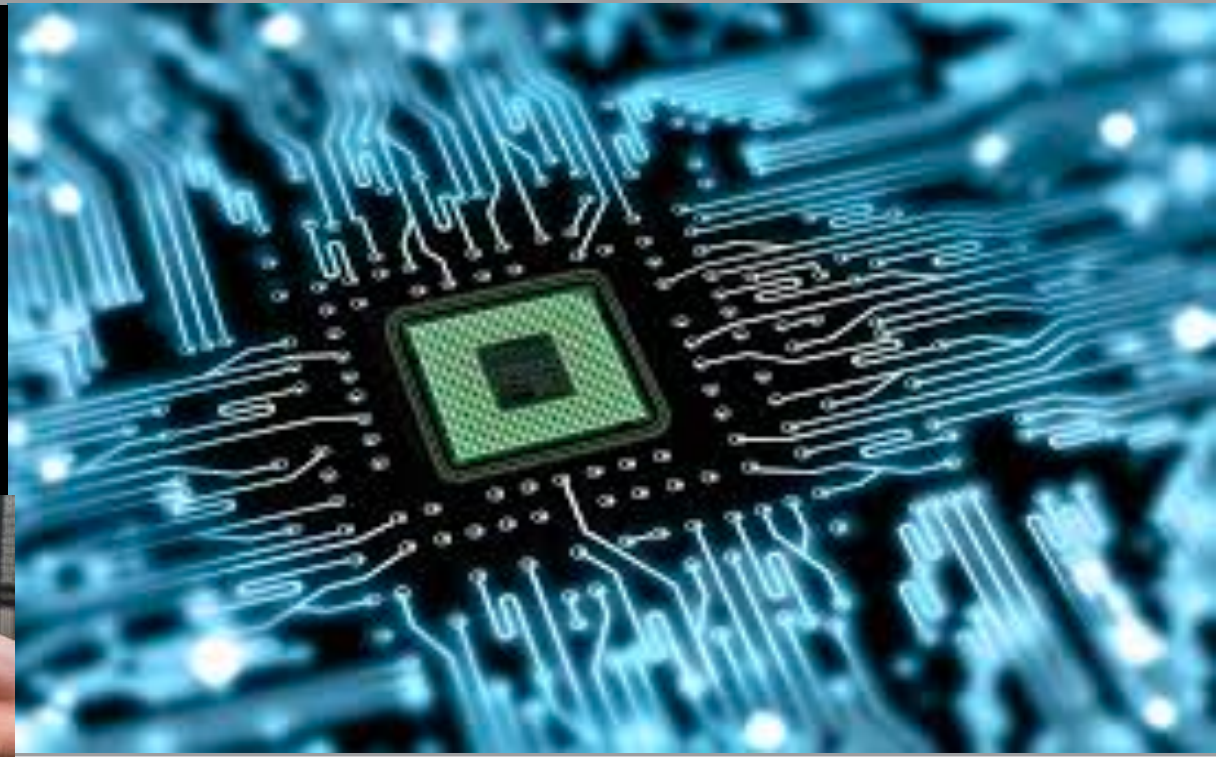


IEEE

Crash Course in Deep Learning

NUI Galway
OÉ Gaillimh

Center for Cognitive, Connected and Computational Imaging
College of Engineering, Science & Informatics, NUI Galway

Cognitive, Connected & Computational Imaging

Outline

- Introduction
- Supervised Learning
  - Convolutional Neural Network
  - Sequence Modelling: RNN and its extensions
- Unsupervised Learning – GANS, Clustering,
- Reinforcement Learning

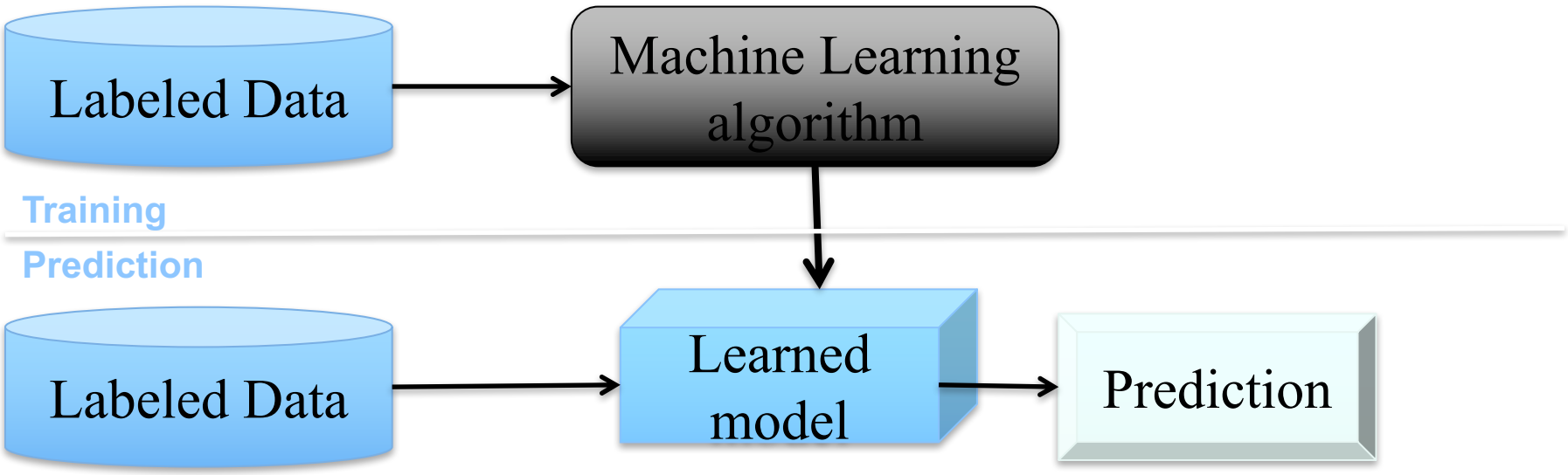
Introduction

Traditional pattern recognition models use hand-crafted features and relatively simple trainable classifier. In deep learning we learn the features instead.

- Manually designed features are often over-specified, incomplete and take a long time to design and validate
- Learned Features are easy to adapt, fast to learn
- Deep learning provides a very flexible, (almost?) universal, learnable framework for representing world, visual and linguistic information.
- Can learn both unsupervised and supervised

Machine Learning Basics

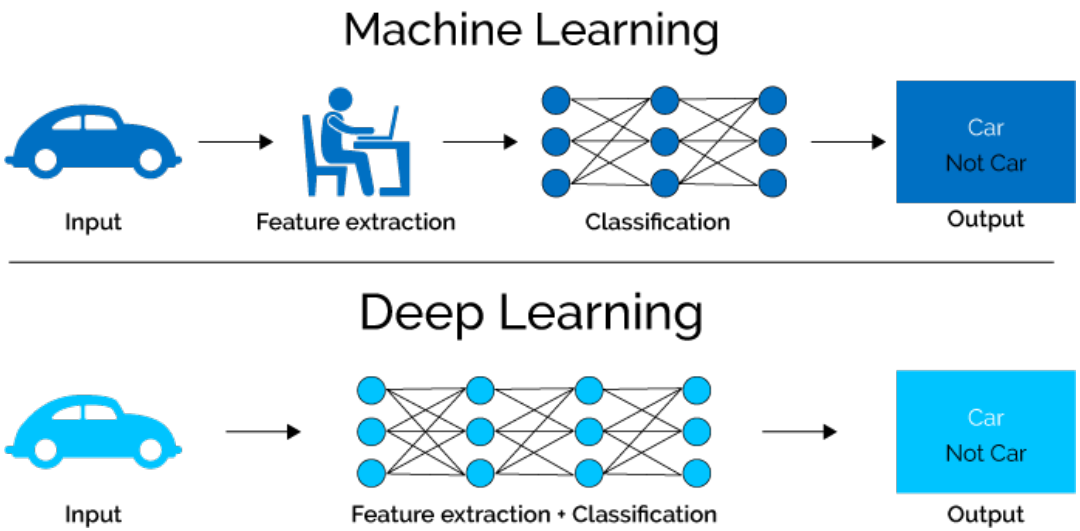
Machine learning gives computers the ability to **learn without being explicitly programmed**



Methods that can learn from and make predictions on data

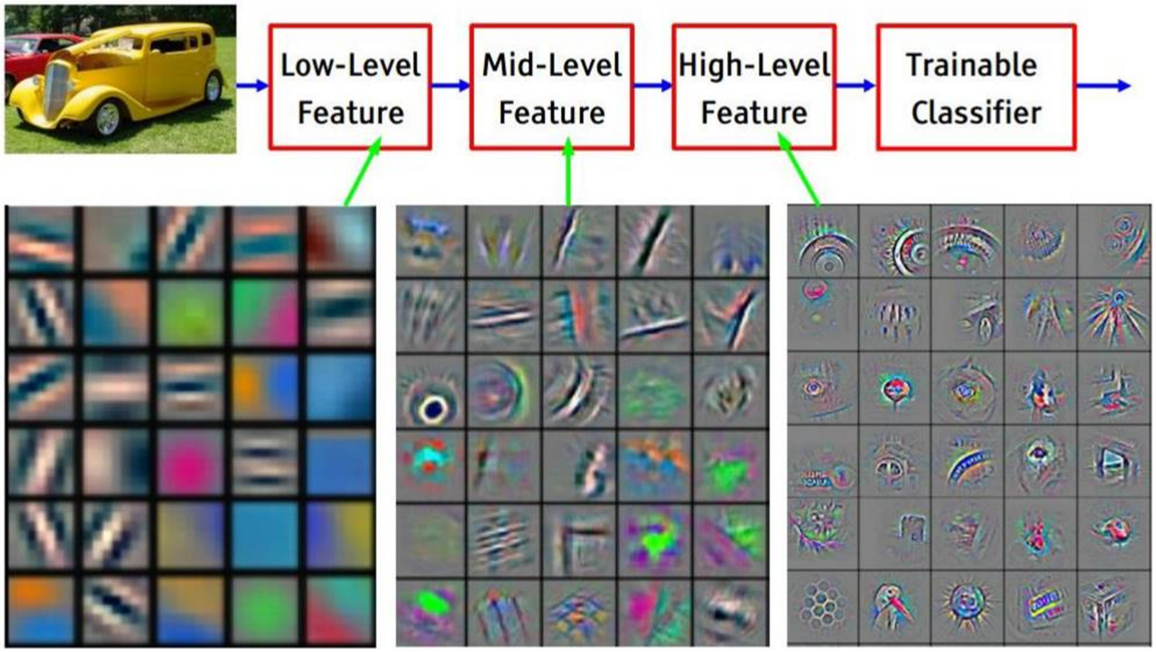
What is Deep Learning (DL) ?

Deep learning algorithms attempt to learn (multiple levels of) representation by using a hierarchy of multiple layers
If you provide the system enough information, it begins to understand it and respond in useful ways.

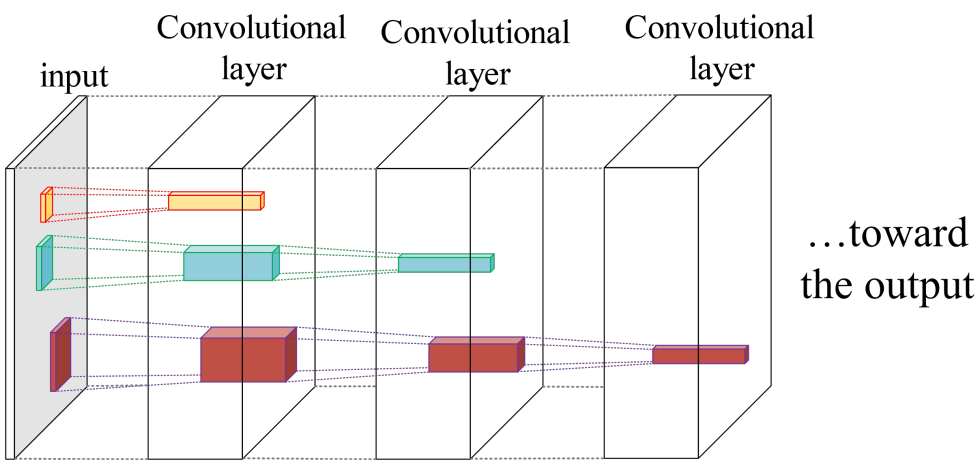


Deep Learning

Deep learning learns rich hierarchical representations (i.e. features) automatically through multiple stage of feature learning process.



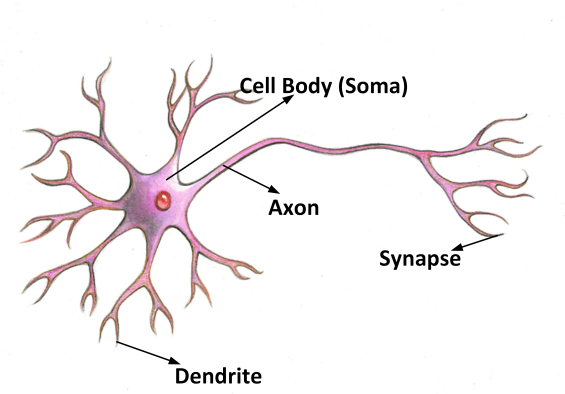
The Mammalian Visual Cortex is Hierarchical

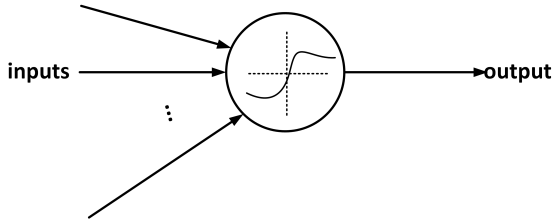


Each module in Deep Learning transforms its input representation into a higher-level one, in a way similar to human cortex.

(Van Essen and Gallant, 1994)

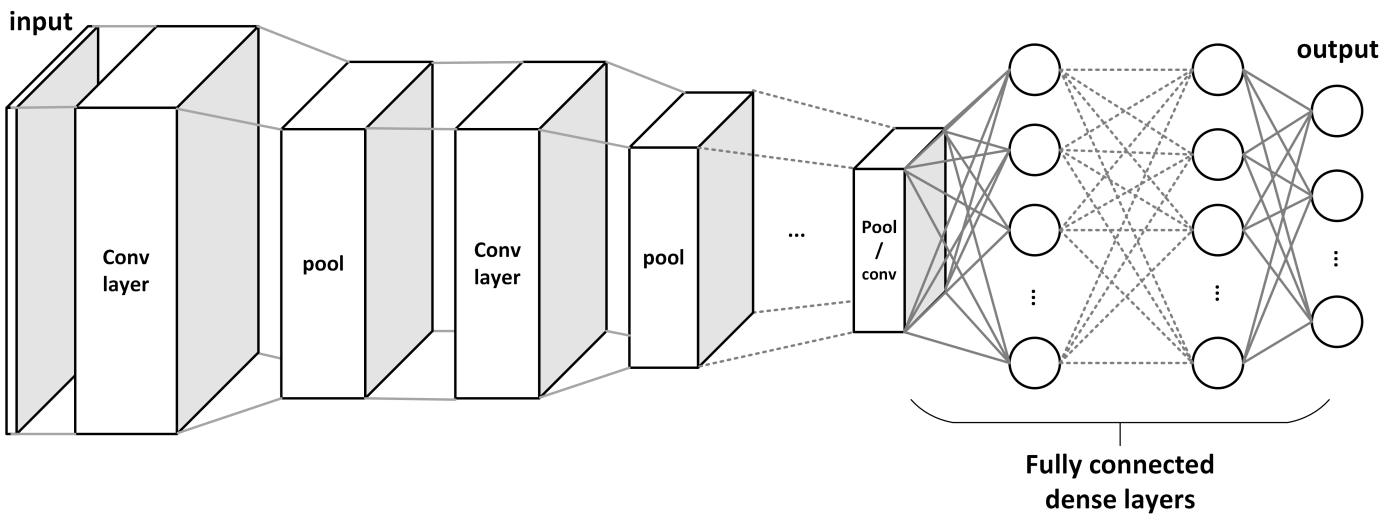
The biological analogy



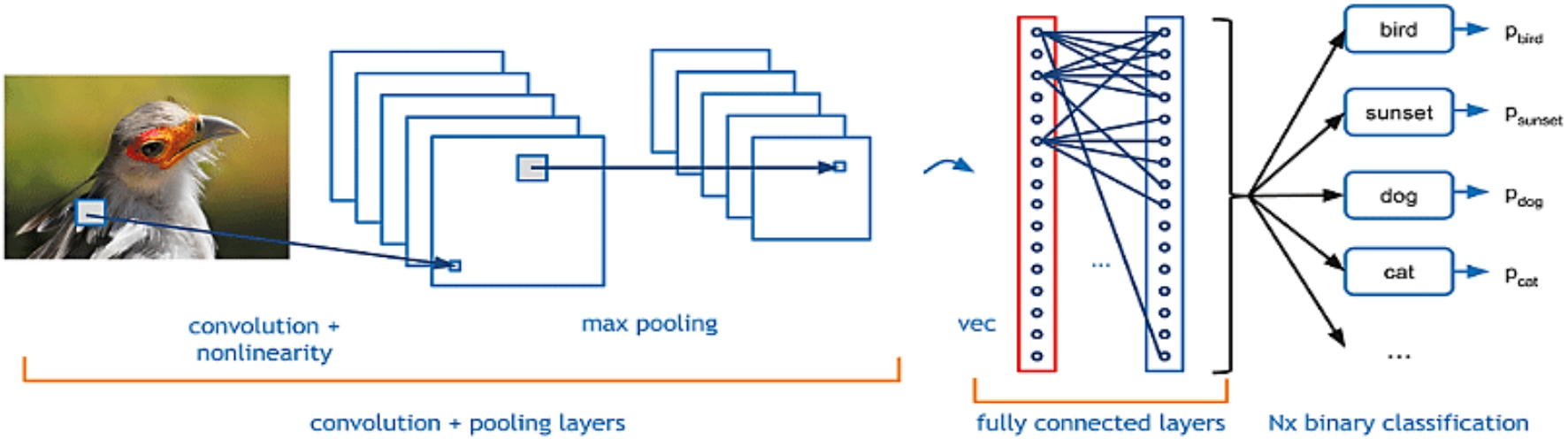


At a very low level, a single unit in a layer is a very simple model of a biological neuron.

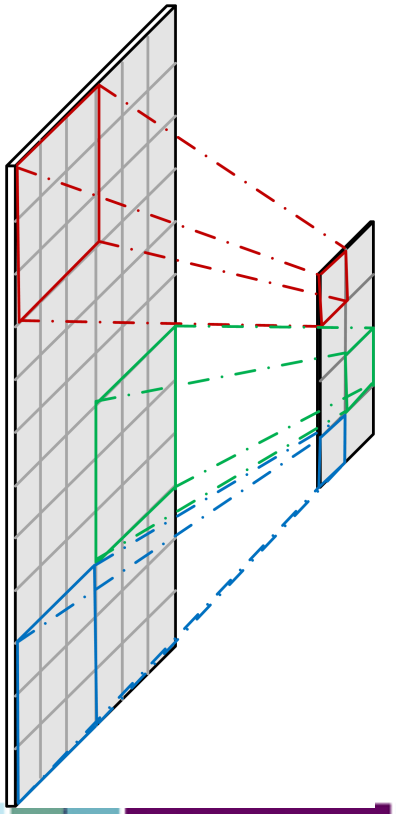
Typical CNN for classification



Deep CNN for Image Classification



Pooling and unpooling



1	5
2	9

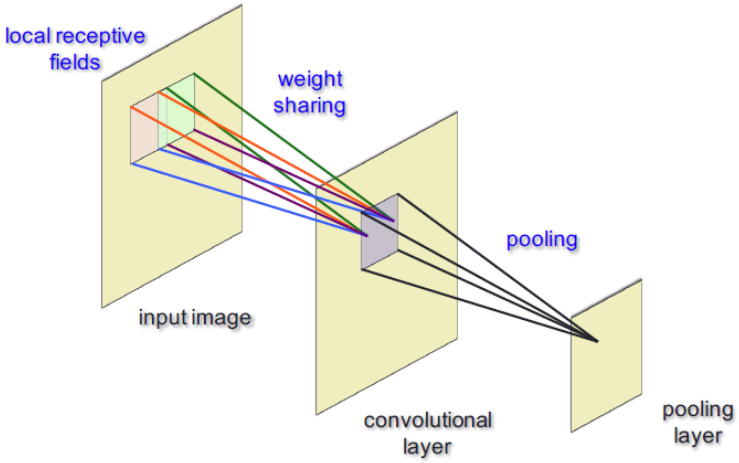
Unpool
2x2

1	1	5	5
1	1	5	5
2	2	9	9
2	2	9	9

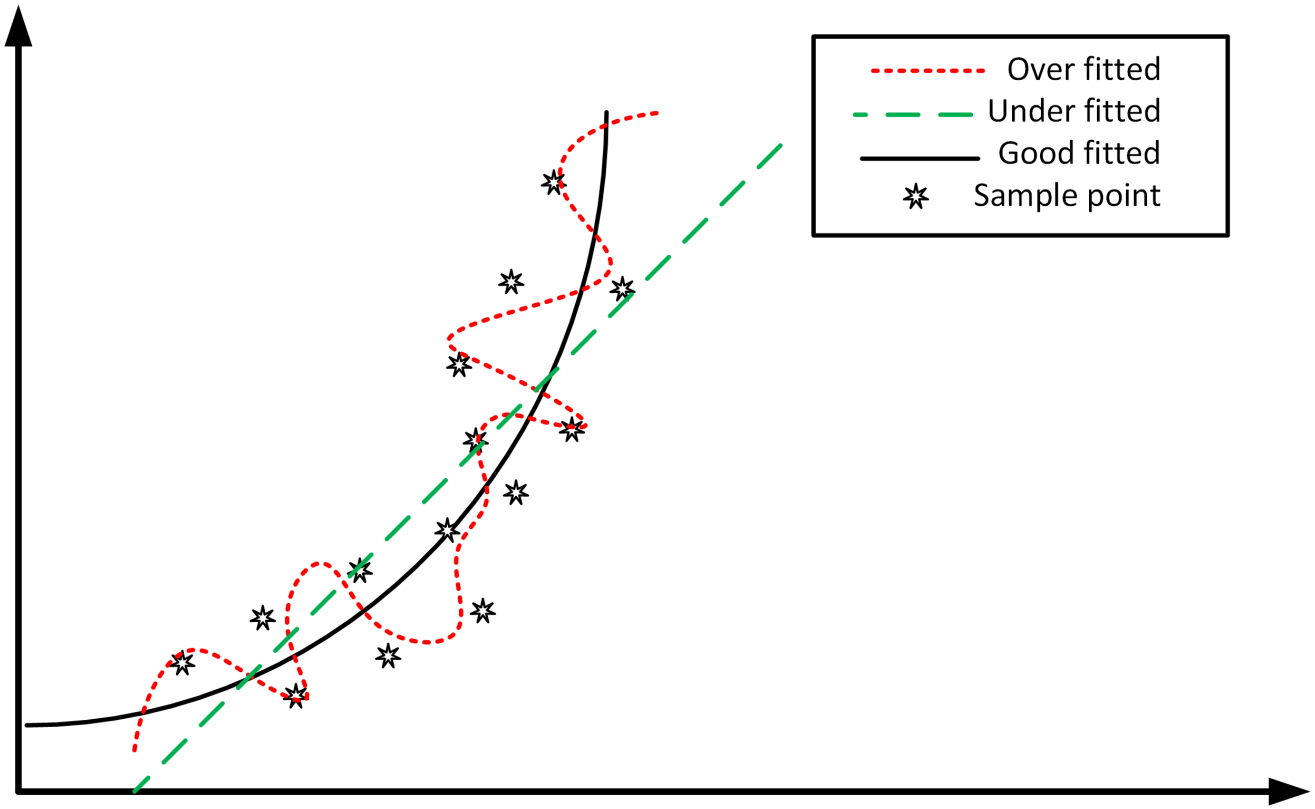
Pooling

Common pooling operations:

- Max pooling: reports the maximum output within a rectangular neighborhood.
- Average pooling: reports the average output of a rectangular neighborhood (possibly weighted by the distance from the central pixel).

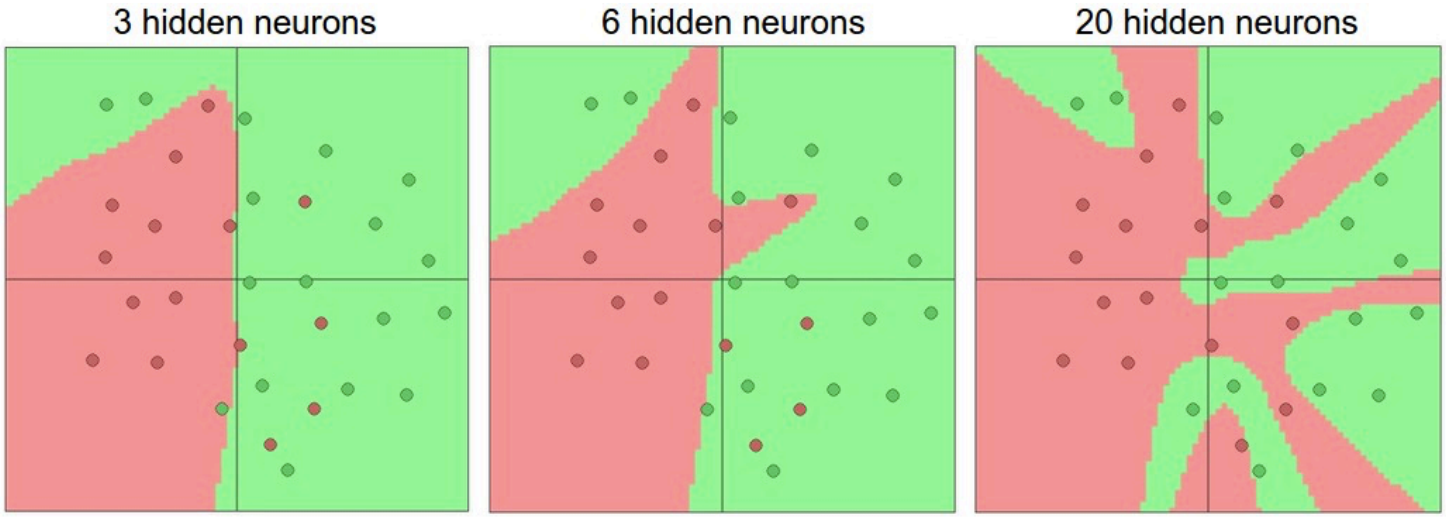


Overfitting and underfitting

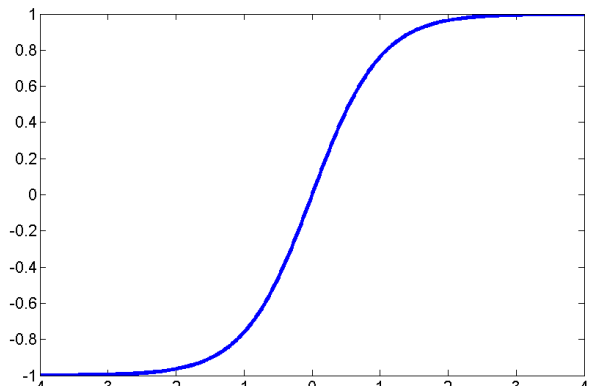


Activation functions

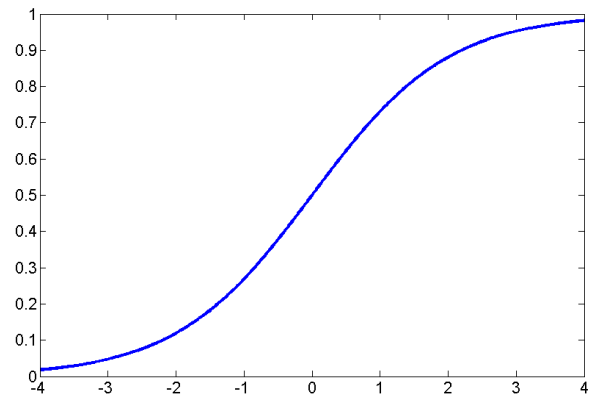
Non-linearities needed to learn complex (non-linear) representations of data, otherwise the NN would be just a linear function $W_1 W_2 x = Wx$



Non-linearity



$f = tanh$



$f = sigmoid$

- Used to be popular in shallow neural networks
- Currently are used in the last layers of the deep neural networks to constrain the output value between [-1 ,1] or [0,1]

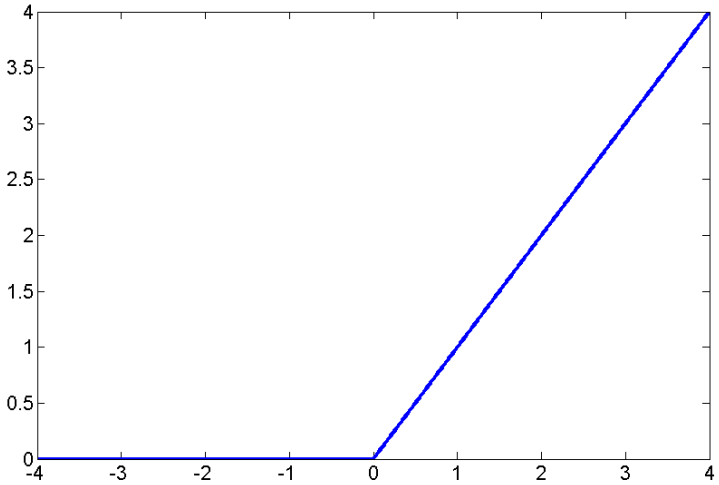
Non-linearity (ReLU)

Sparse activation.

Better gradient propagation.

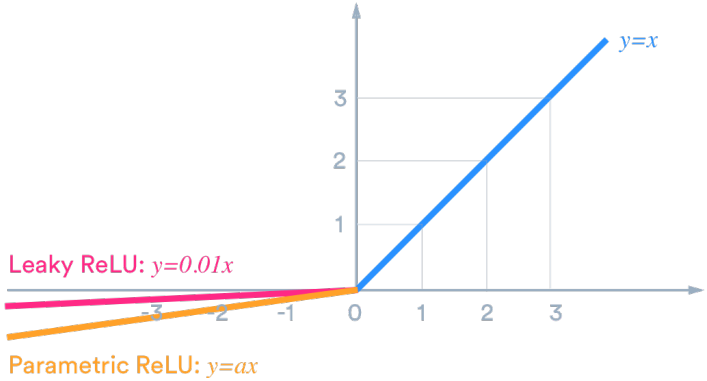
Simplicity in implementation.

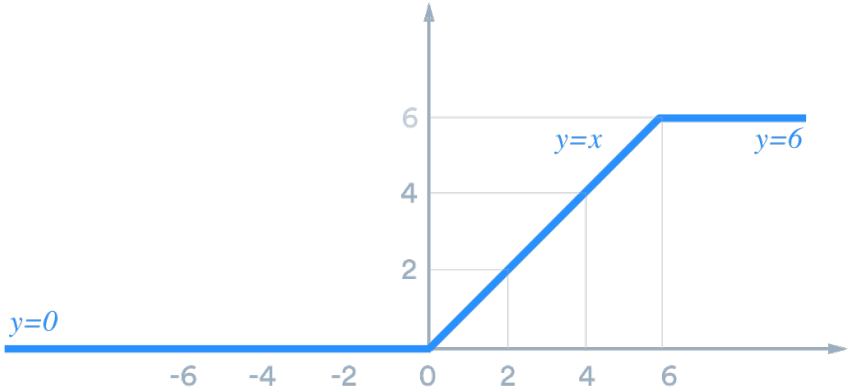
The nonlinearity is around 0 where the data tend to be.

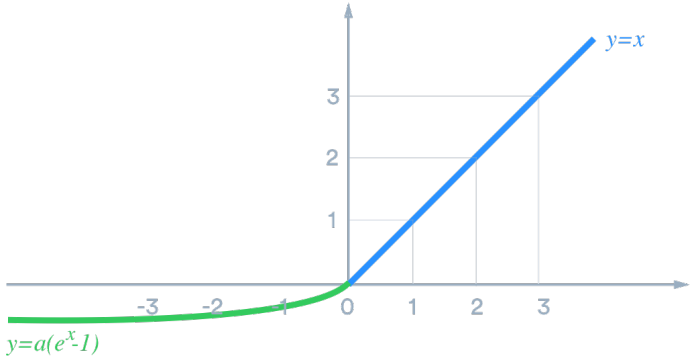


R. H. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit," Nature, vol. 405, no. 6789, p. 947, 2000

Non-linearity (ReLU)



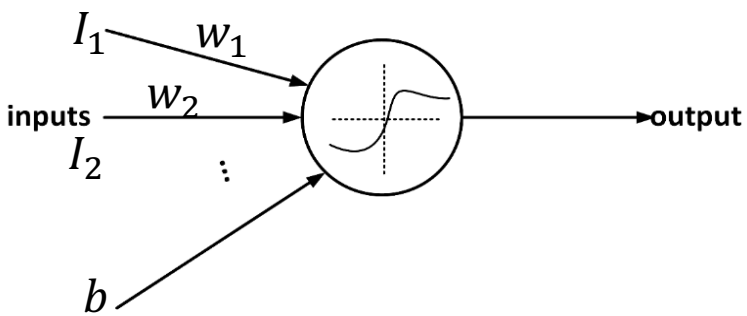




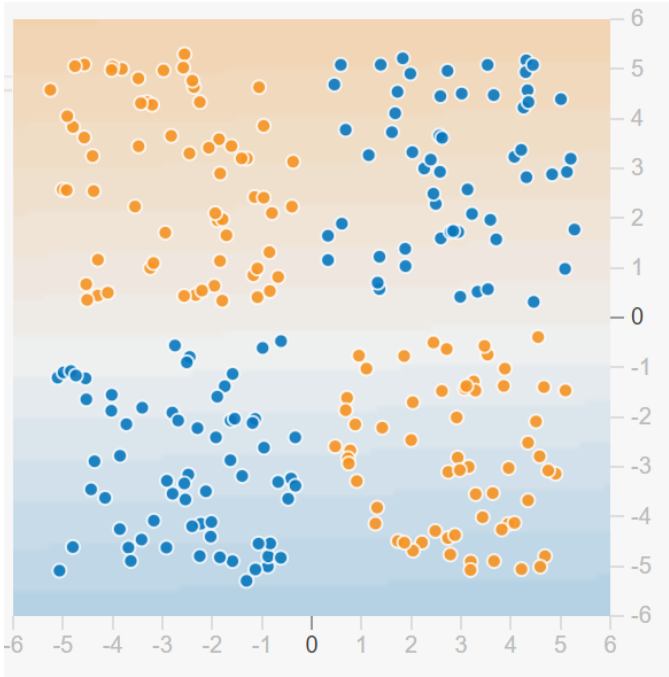
Graphs from: https://www.tinymind.com/learn/terms/relu

Neural Networks

- Neuron

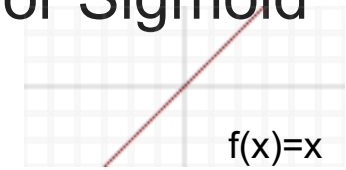


$$O = f\left(\sum_i w_i I_i + b\right)$$



https://playground.tensorflow.org

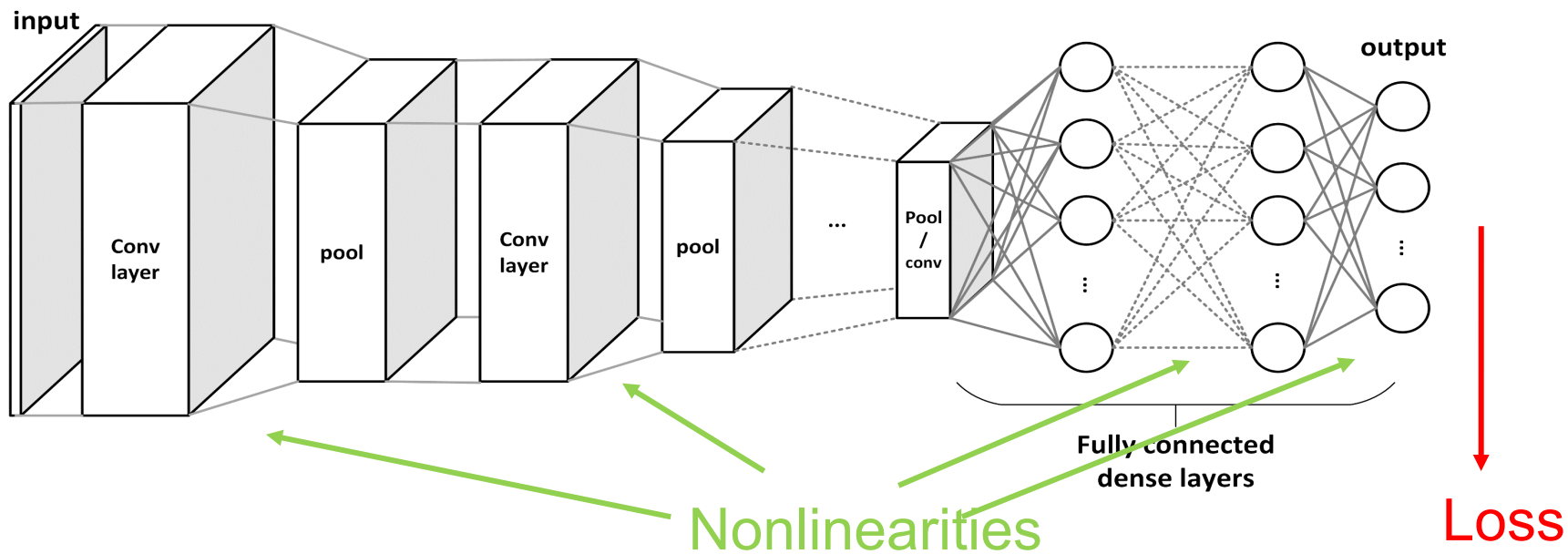
Loss functions and output

	Classification	Regression
Training examples	R^n x {class_1, ..., class_n} (one-hot encoding)	R^n x R^m
Output Layer	Soft-max [map R^n to a probability distribution] $P(y=j \mid \mathbf{x}) = \frac{e^{\mathbf{x}^\top \mathbf{w}_j}}{\sum_{k=1}^{K} e^{\mathbf{x}^\top \mathbf{w}_k}}$	Linear (Identity) or Sigmoid 
	Cross-entropy	Mean Squared Error

$$J(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\left[y_k^{(i)} \log \hat{y}_k^{(i)} + \left(1 - y_k^{(i)}\right) \log\left(1 - \hat{y}_k^{(i)}\right)\right]$$

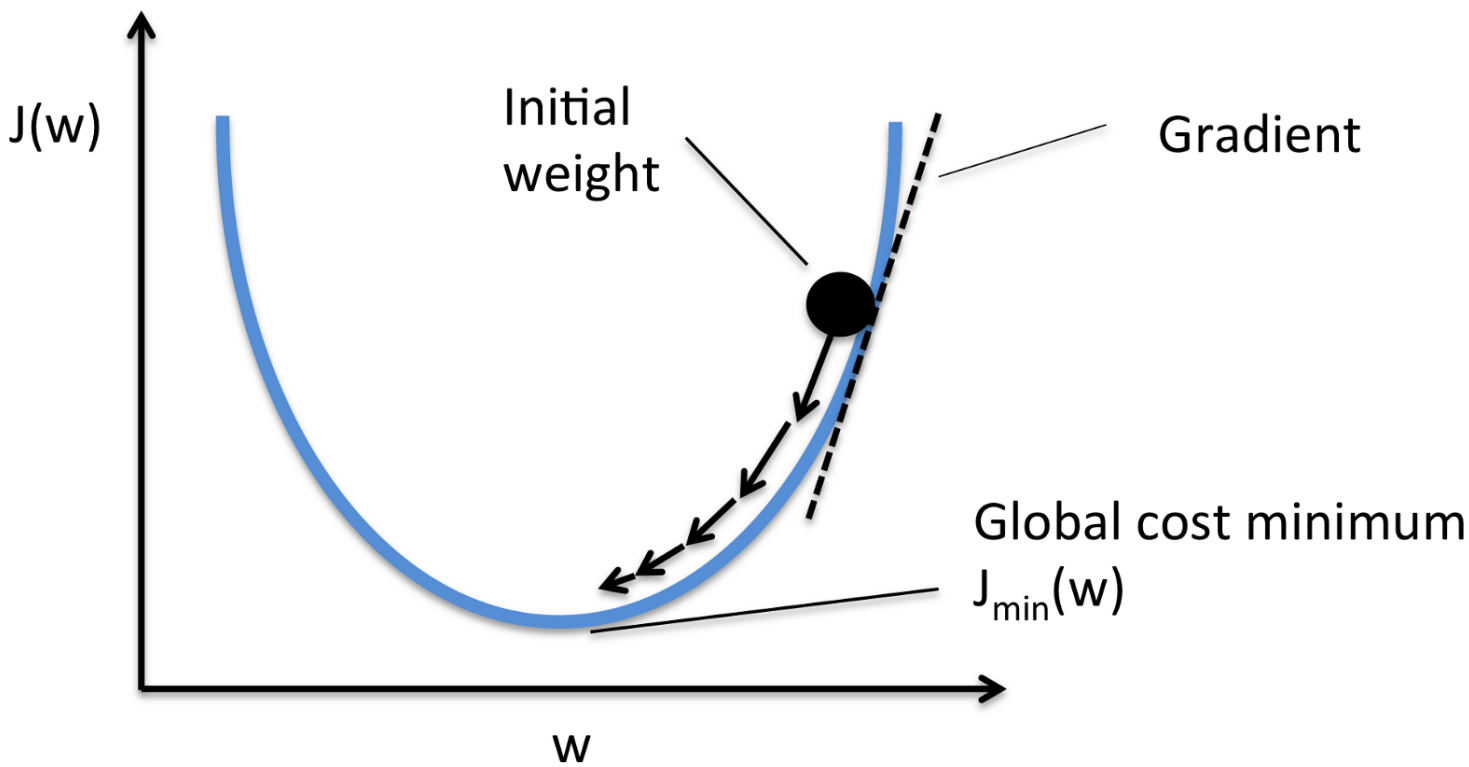
$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - \hat{y}^{(i)}\right)^2$$

(training -backpropagation-)

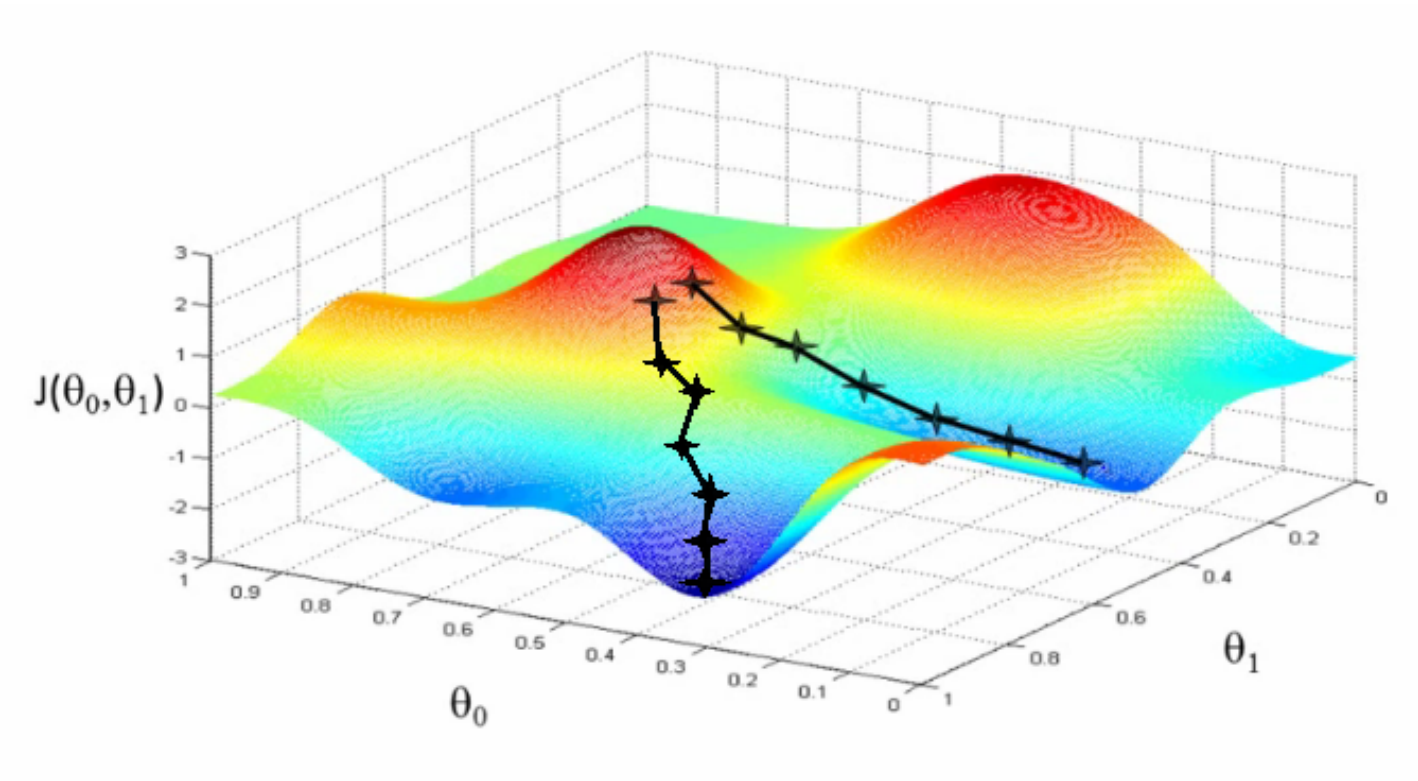


Backpropagation

Backpropagation- Gradient Descent (GD)



Backpropagation- Gradient Descent (GD)

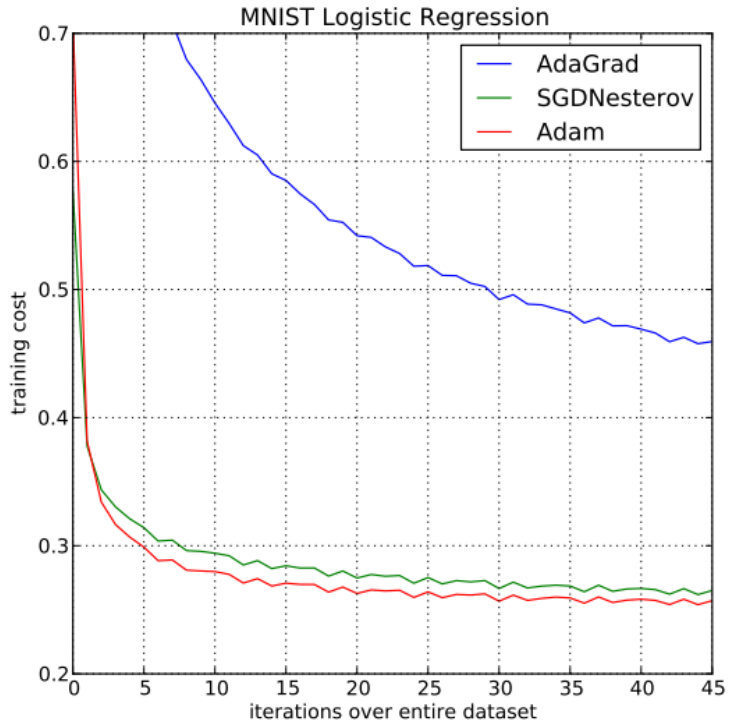


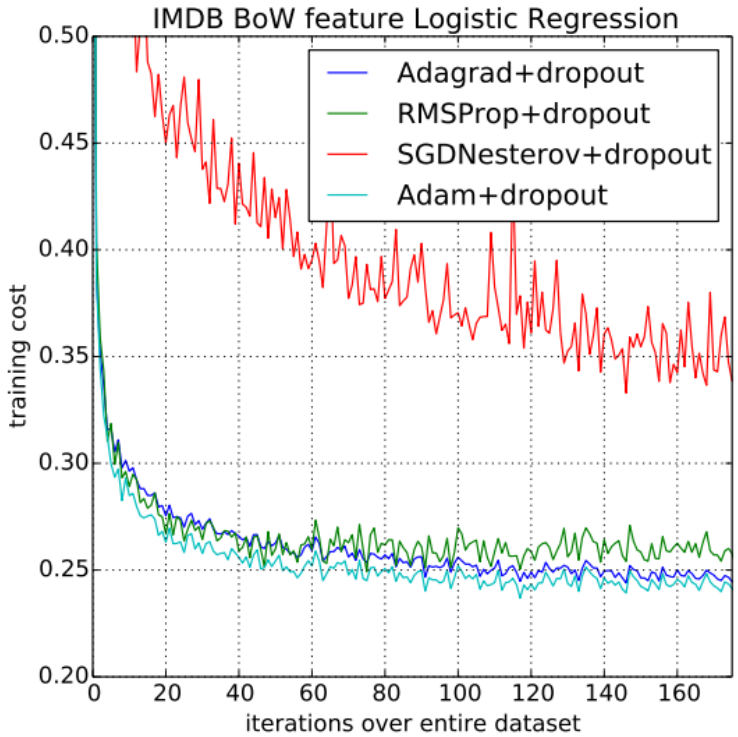
https://medium.com/ai-society/hello-gradient-descent-ef74434bdfa5

Optimization methods

- Is probably the most sophisticated part of training a network.
- Gradient Descent is the most popular method in optimization.
  - Stochastic Gradient Descent (SGD)
  - SGD with Nestrov momentum
  - AdaGrad
  - AdaDelta
  - RMSprop
  - Adaptive Moment Estimation (ADAM)

Optimization methods





Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Sequence Modelling

- Why do we need RNN?
- What are RNNs?
- RNN Extensions
- What can RNNs can do?

Why do we need RNNs?

The limitations of the Neural network (CNNs)

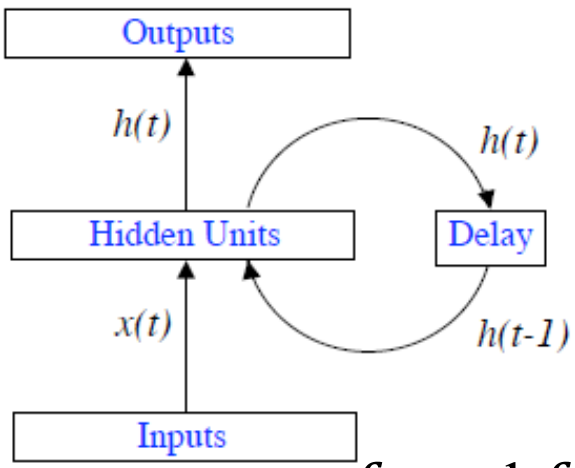
- Rely on the assumption of independence among the (training and test) examples.
  - After each data point is processed, the entire state of the network is lost
- Rely on examples being vectors of fixed length

We need to model the data with temporal or sequential structures and varying length of inputs and outputs

- Frames from video
- Snippets of audio
- Words pulled from sentences

What are RNNs?

Recurrent neural networks (RNNs) have the ability to selectively pass information across sequence steps, while processing sequential data one element at a time.



Allow a 'memory' of previous inputs to persist in the network's internal state, and thereby influence the network output

$$h(t) = f_H(W_{IH}x(t) + W_{HH}h(t-1))$$

$$y(t) = f_O(W_{HO}h(t))$$

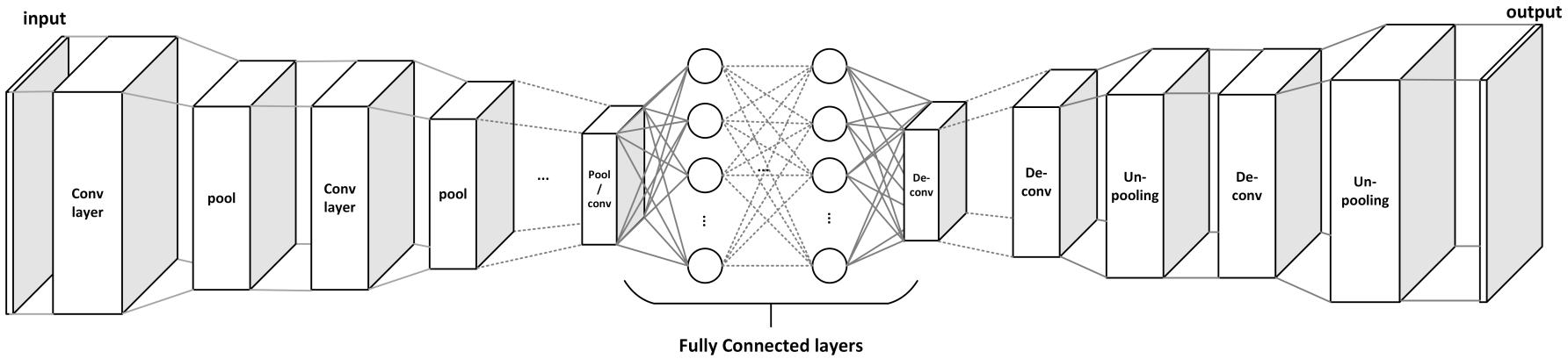
f_H and f_O are the activation function for hidden and output unit; W_{IH}, W_{HH}, and W_{HO} are connection weight matrices which are learnt by training

Unsupervised Learning

- Autoencoders
- Deep Autoencoders
- Denoising Autoencoders
- Stacked Denoising Autoencoders
- GAN
- Artistic style transfer

Autoencoders

An autoencoder learns to compress data from the input layer into a short code, and then uncompress that code into something that closely matches the original data.



Deep Learning

- Deep learning technologies are already widely deployed.
- In your phone DL is used for speech recognition, image recognition, image processing, intelligent search.
- Deep learning models are embedded in APIs which you can call without even knowing you are using deep learning.
- Intelligent devices are the near future.