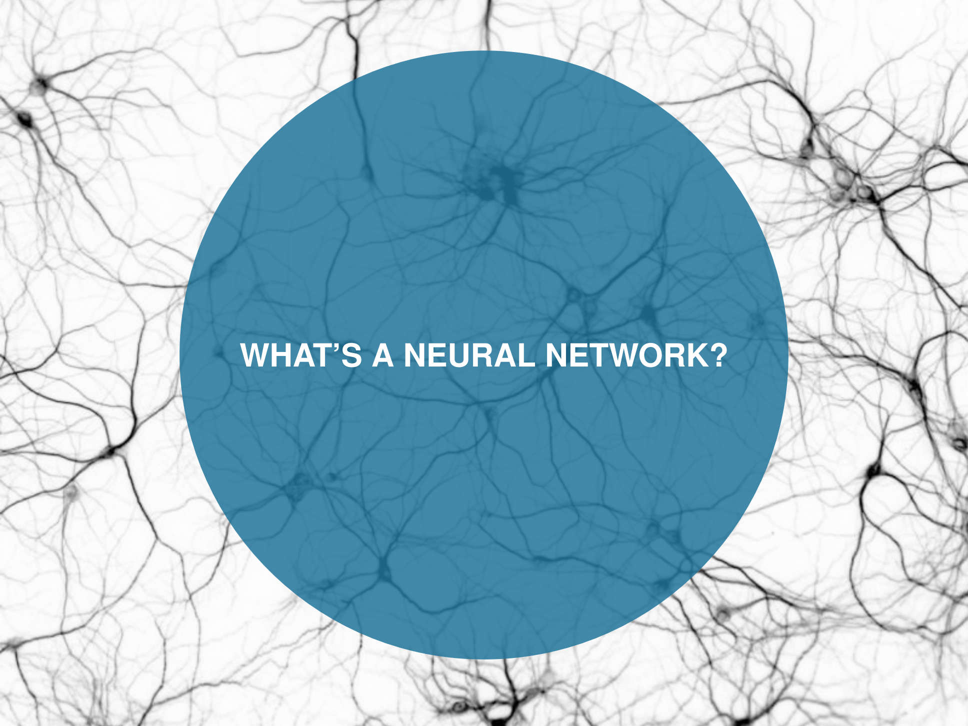
Artifical neural networks and deep learning

James D Pearce • University of Victoria • jdpearce@uvic.ca



WHAT’S A NEURAL NETWORK?

- Originally models of biological neural networks
- Now based in a now based in a probabilistic framework
- Example of what is known as a machine learning algorithm
- used for classification, regression and feature discovery

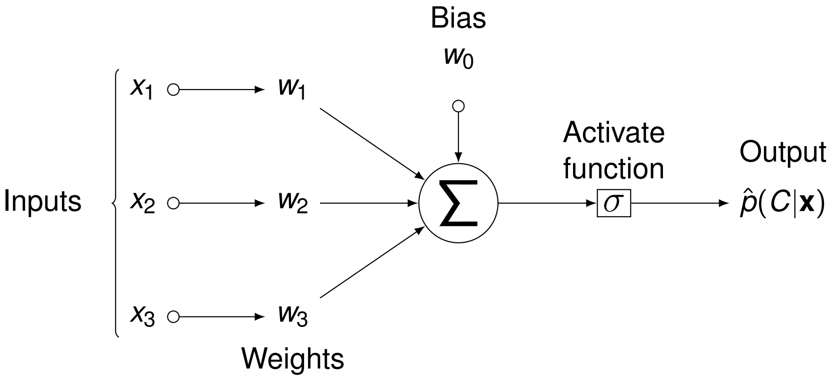
Machine learning: "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*"

- Mitchell, T. (1997). Machine Learning, McGraw Hill

Example:

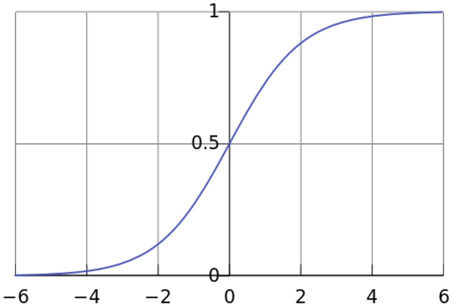
- *T*: classify handwritten digits 0-9
- *E*: Data set of handwritten digit images with associated human generated labels 0-9
- *P*: Accuracy of classification with respect to associated labels

The logistic unit is the basic building block for neural networks.

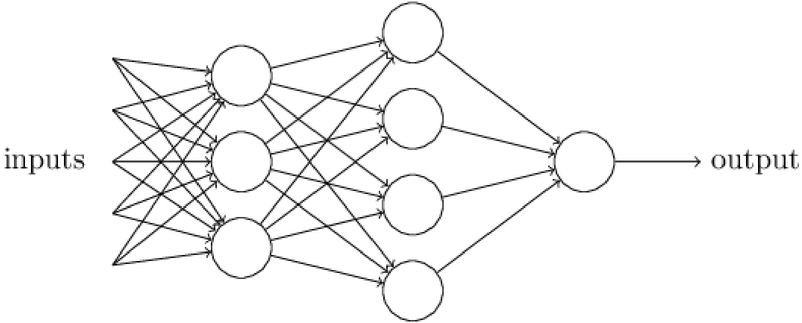


The activation is a pdf-like function of a linear combination of the inputs:

$$\sigma(\mathbf{x}) = \frac{1}{1 + \exp(-\sum x_i w_i + w_0)} \approx \hat{p}(C|\mathbf{x}, \mathbf{w}) \tag{1}$$



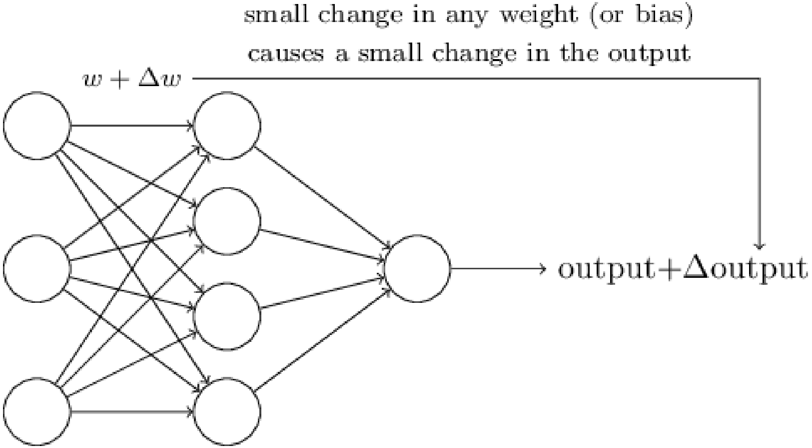
These logistic untis can be stacked in layers, where the output of the first layer becomes an input to the next. This is called a neural network.



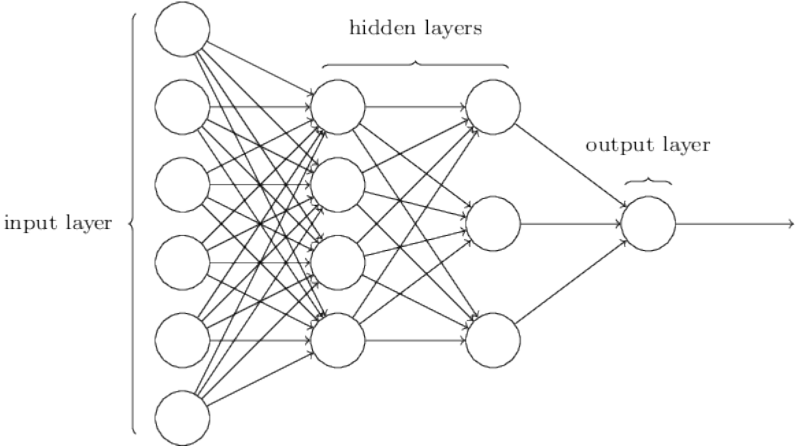
Training a neural net consists of choosing the weights, **w**, such that the error rate is minimized. This equivalent to finding the maximum likelihood estimator for the weights:

$$\hat{\mathbf{w}} = \arg\max_{\mathbf{w}} \hat{p}(\mathbf{x}|\mathbf{w}) \tag{2}$$

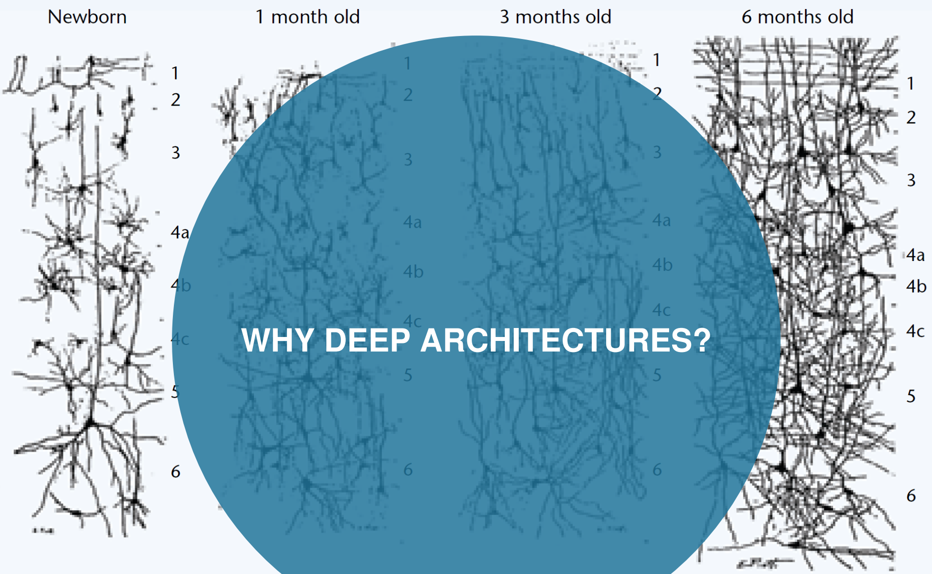
Training is done by iteratively updating the weights, usually using some variant of gradient decent.



When more than one hidden layer is present the model is *deep*

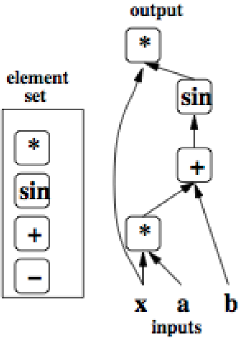


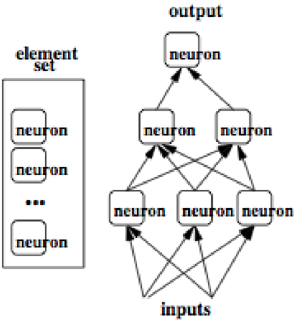
WHY DEEP ARCHITECTURES?



Golgi stain preparations (from Conel, 1939–67)

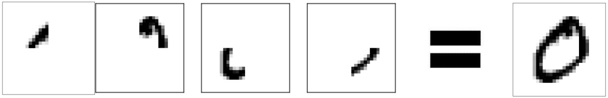
Calculating $x \sin(ax + b)$ should be done as a sequence of operations, although one could take the Fourier series of the function and calculated it all in one sum, but this is not efficient!





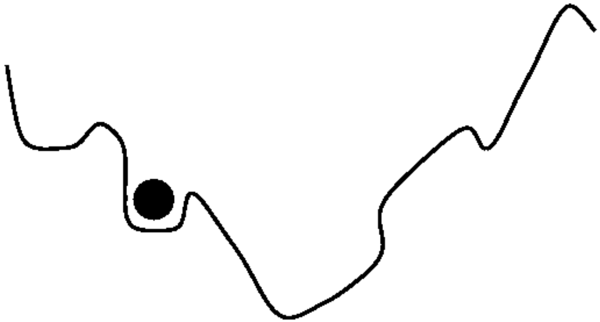
But more importantly beep architectures allow for higher level features to be built up out of a shared set of lower level features.

Example: 0 is made up lower level features (pen strokes)

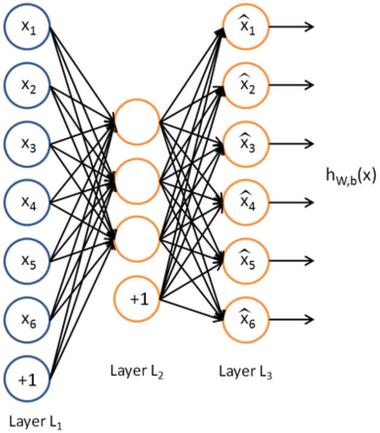


However there are still some outstanding issue with training deep nets:

- Vanishing gradient problem
- Local minimum problem

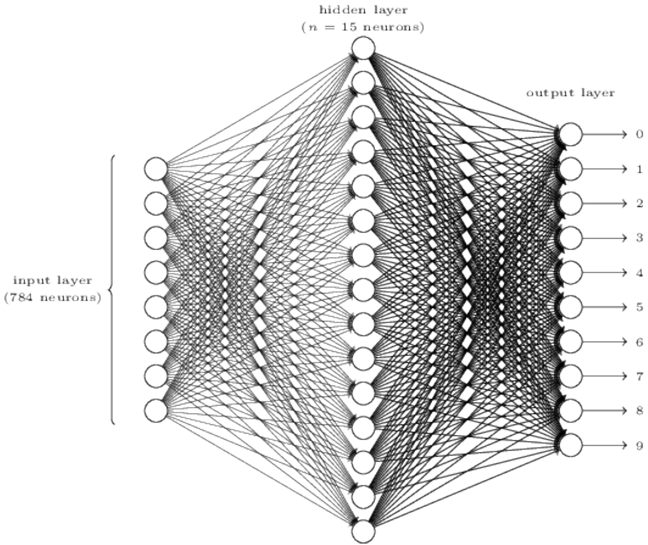


Auto-encoders can be used as a pre-training method. This helps to solve the local minimum problem.

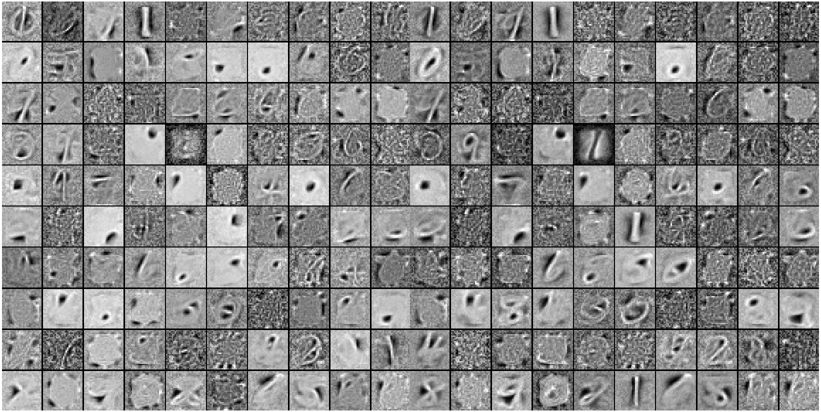



HANDWRITTEN DIGIT CLASSIFICATION

Example of the simplest architecture form MNIST data:



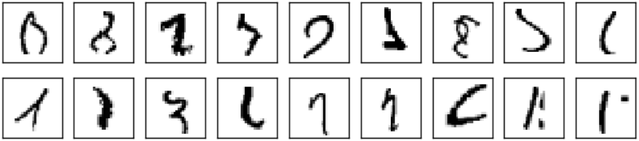
Weight vectors produced by auto-encoders. They learn pen strokes without any supervision



Results of experiments:

Network type	Architecture	Weights	Runtime (min)	Error rate (%)
Logistic	784-10	7,840	0.664	7.48
MLP	784-100-100-100-10	99,400	21.4	2.72
MLP	784-500-10	397,000	72.1	1.67
MLP	784-1000-10	794,000	244.2	1.66
MLP	784-1000-1000-10	1,794,000	579.3	1.46
MLP	784-3500-10	2,779,000	567.5	1.82
MLP	784-1000-1000-1000-10	2,794,000	1,394.9	1.58
SDAE	784-100-100-100-10	99,400	21.4+30.0	2.76
SDAE	784-1000-1000-1000-10	2,794,000	1,336.9+1,183	1.28

Probably not all of the digits can be classified . . .



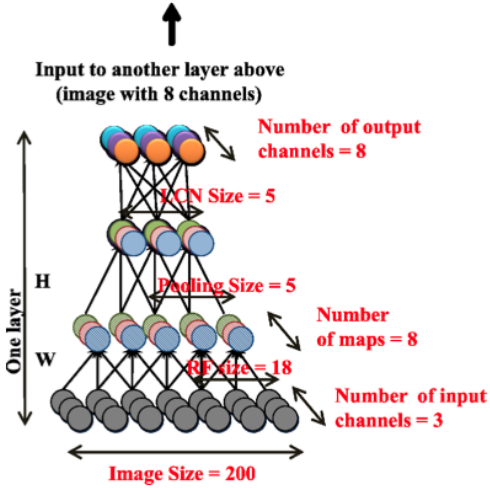
UNSUPERVISED LEARNING

"A typical adult human brain has about 10^{14} synapses (connections), and a typical human lives on the order of 10^9 seconds. Thus, even if each synapse is parameterized by just a one bit parameter, a learning algorithm would require about $10^{14}/10^9 = 10^5$ bits of information per second to"learn" all the connections in the brain. It seems extremely unlikely that this many bits of labeled information are available (say, from a human's parents or teachers in his/her youth)."

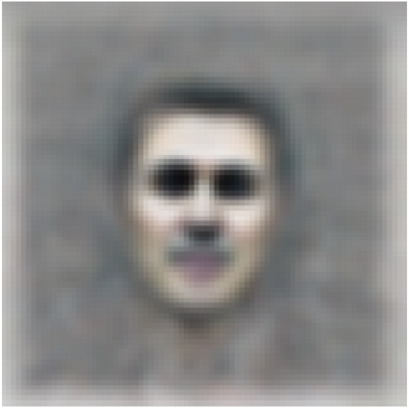
– Geoffrey Hinton, communication

Humans seem to learn almost entirely from unlabelled data.
Can neural nets do the same?

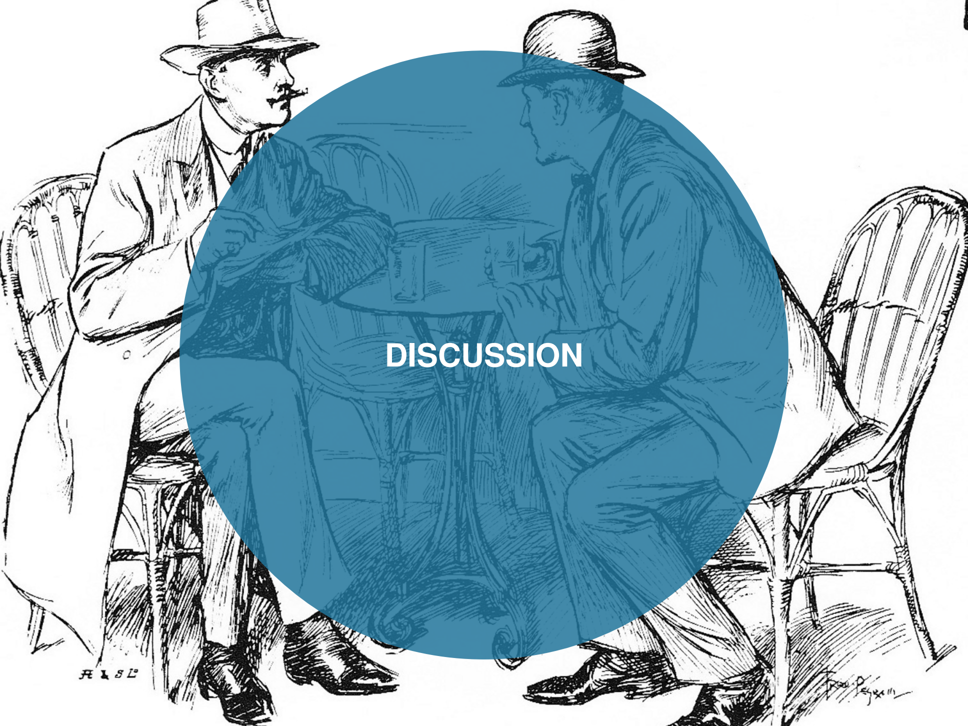
Google trained a 9-layer net with 1 billion parameters using 16,000 cores.



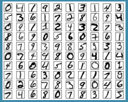
After 3 days of training on youtube images (without labels of any kind) it was shown pictures of human faces and cats. The weight vectors of the maximally excited neurons are shown.

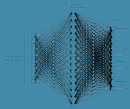


DISCUSSION



SOURCES


yann.lecun.com/exdb/mnist/


neuralnetworksanddeeplearning.com

WIKIPEDIA wikipedia.com