Deep Neural networks

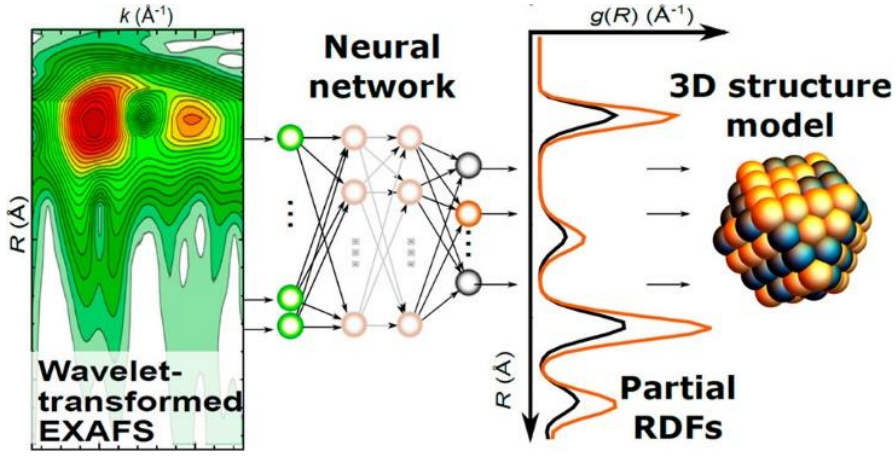
Associated notebook: https://github.com/keeeto/reading-ml-chemistry/blob/master/02_DNN.ipynb

Overview

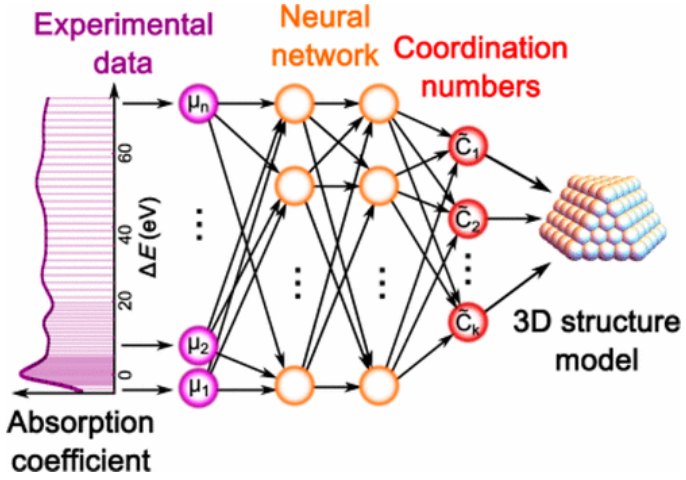
- History of deep neural nets (DNN)
- Layers of a DNN
- Structure of a neuron
- Activation functions
- Backpropagation
- Optimisation
- Regularisation

Showcase

- Multi Layer Perceptrons (MLPs) for interpreting X-ray spectroscopy
- Transforming an EXAFS signal into a partial RDF to assist with 3D model building using an MLP
- XANES spectra used to estimate coordination numbers of atomic sites using an MLP
- Processing of each spectrum takes less than 1 second
- Gets coordination up to the 4th coordination shell



Nano Lett. 2019, 19, 1, 520–52

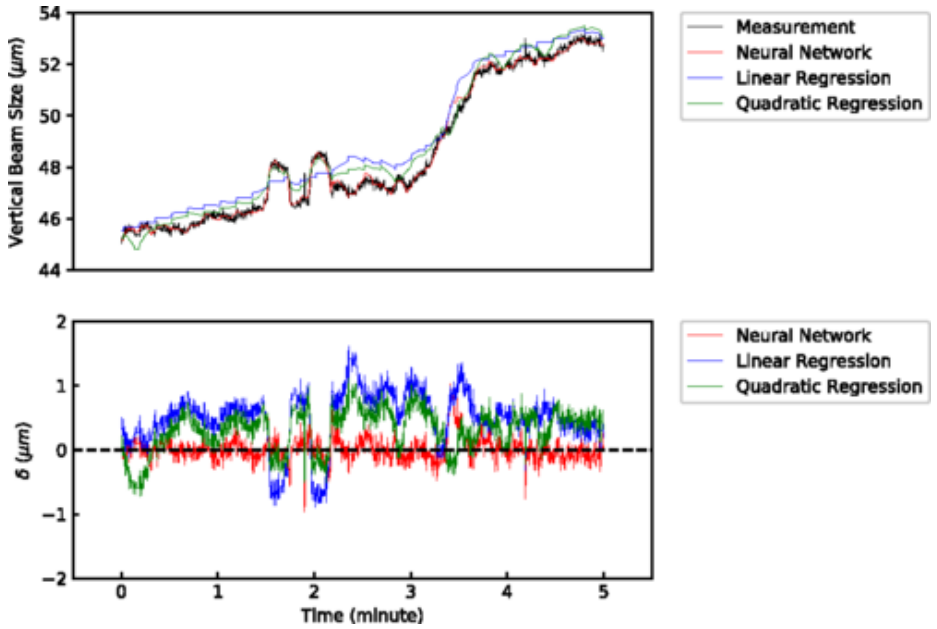


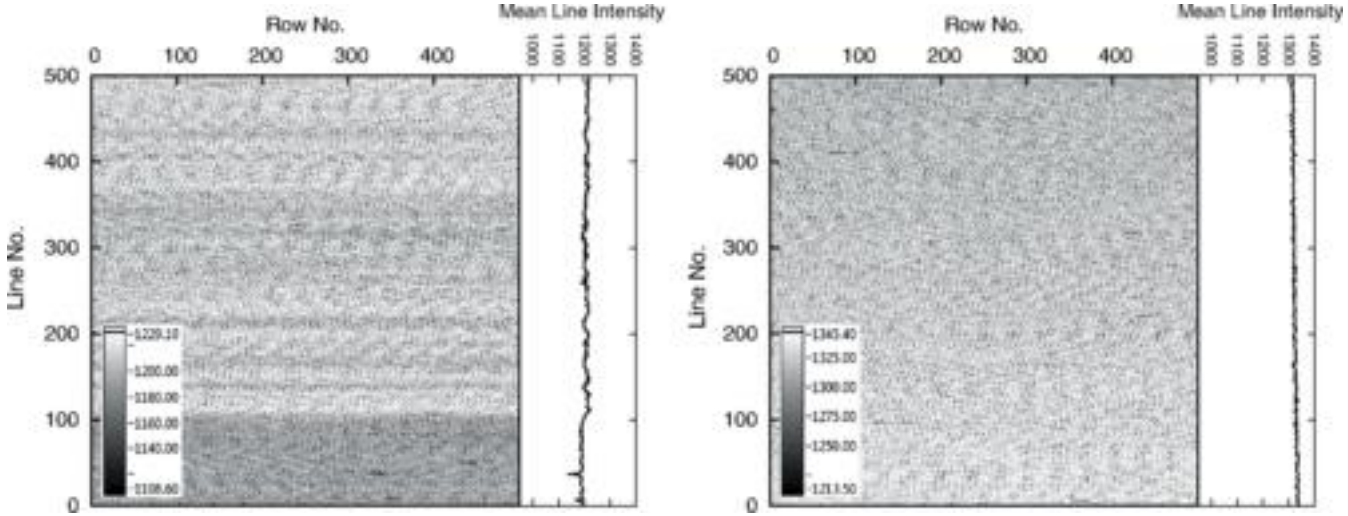
J. Phys. Chem. Lett. 2017, 8, 20, 5091–5098

Science and Technology Facilities Council

Showcase

- Stabilisation of synchrotron beam at Advanced Light Source
- “Learned” how changes in the configurations and positions of the magnetic excitations affected the width of the resulting photon beam
- Precision is better than the 2–3% previously achieved at the ALS
- Allows the synchrotron to probe the dynamics of chemical reactions





Phys. Rev. Lett. 2019, 123, 194801

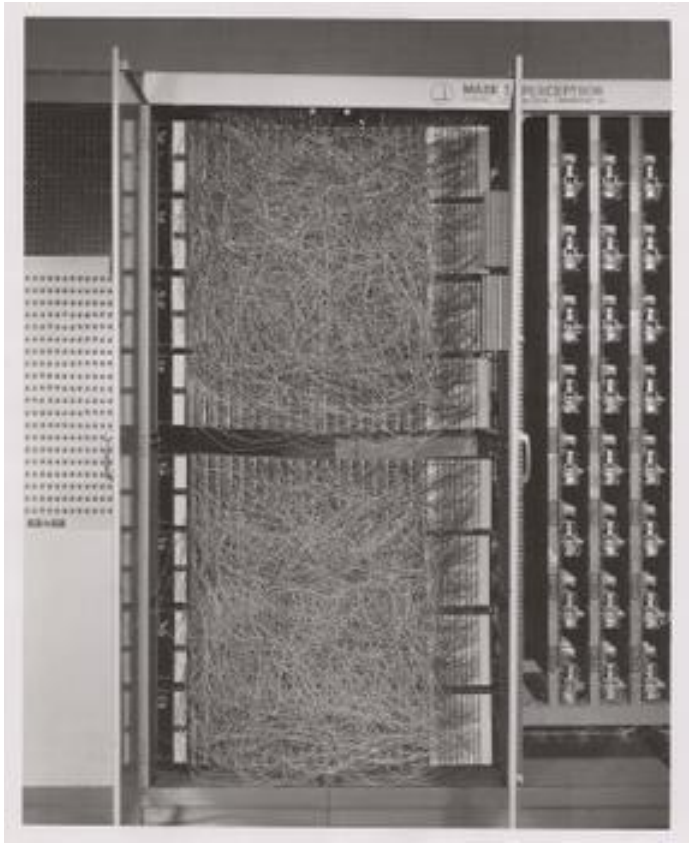
Science and Technology Facilities Council

Neural networks

- Early NN
- Originally a device
- Intended for binary classification

- Produces a single output from a matrix of inputs, weights and biases

$$y = \phi(\sum_i w_i x_i + b) = \phi(\mathbf{w}^T \mathbf{x} + b)$$

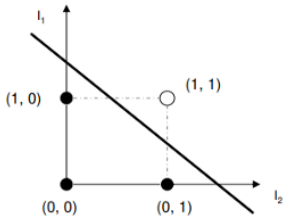


Science and Technology Facilities Council

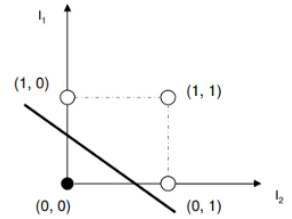
Neural networks

- Single layer
- Minsky and Papert showed they could not solve non-linear classification

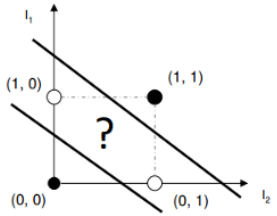
AND		
I_1	I_2	out
0	0	0
0	1	0
1	0	0
1	1	1

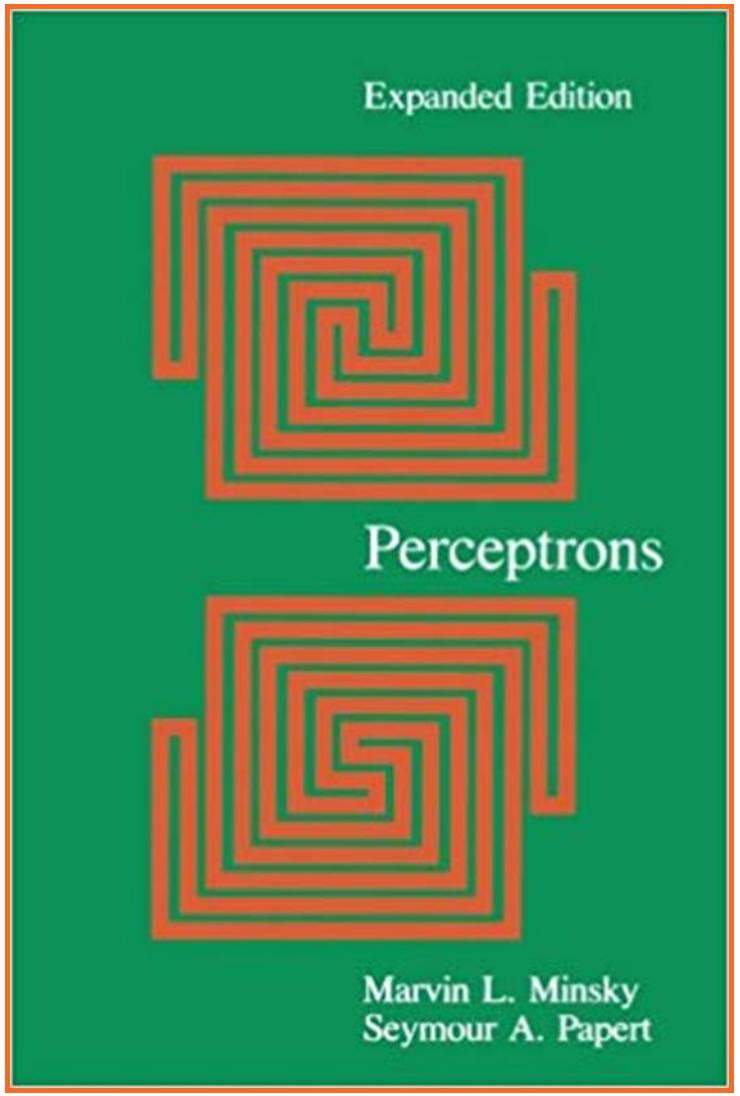


OR		
I_1	I_2	out
0	0	0
0	1	1
1	0	1
1	1	1



XOR		
I_1	I_2	out
0	0	0
0	1	1
1	0	1
1	1	0

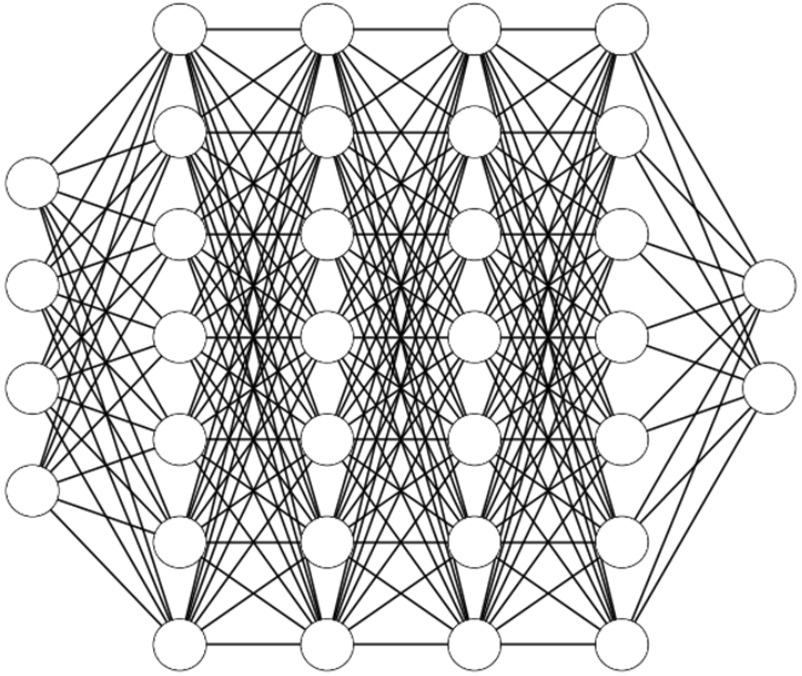




Science and Technology Facilities Council

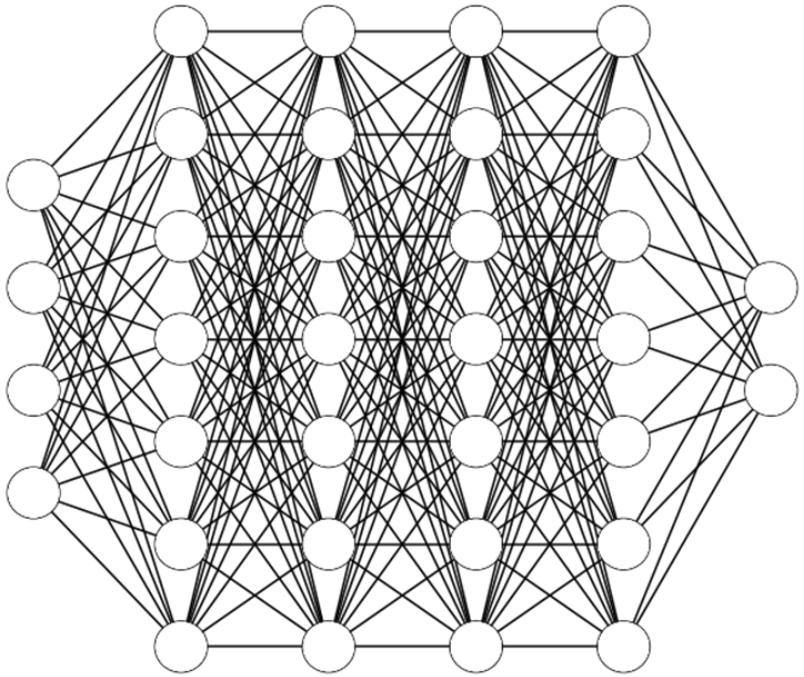
Neural networks

- Back propagation
- Now gradients could be used to minimise error
- Modifications back propagate through the network using the chain rule



Deep neural networks: Multi layer perceptron

Input layer



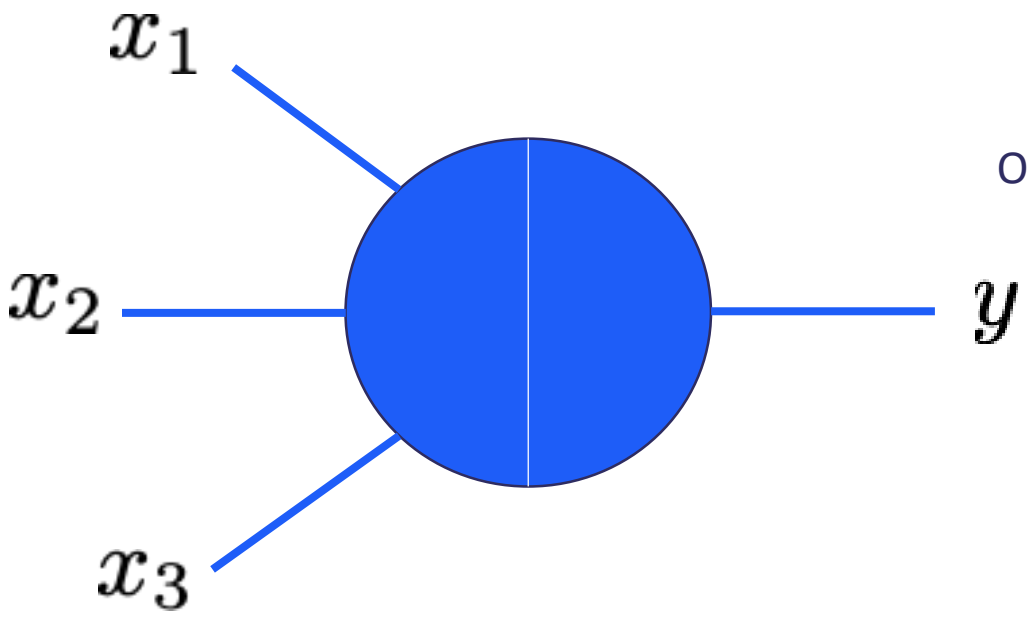
Output layer

Hidden layers

Science and Technology Facilities Council

Dense layers

- Also called fully connected layers



Activation function

Output signal — $y = g(z)$

$$z = \boldsymbol{w}^T.\boldsymbol{x} + b$$

Bias

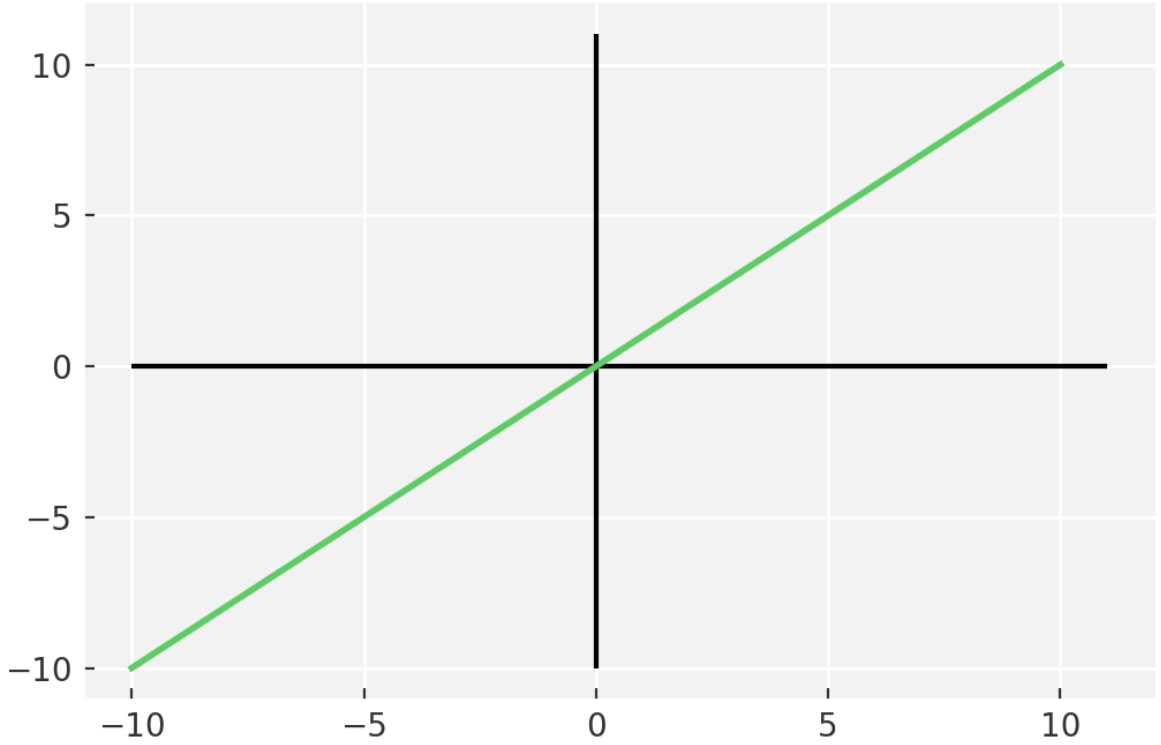
Weights

Input signals

Science and Technology Facilities Council

Activation function: Linear

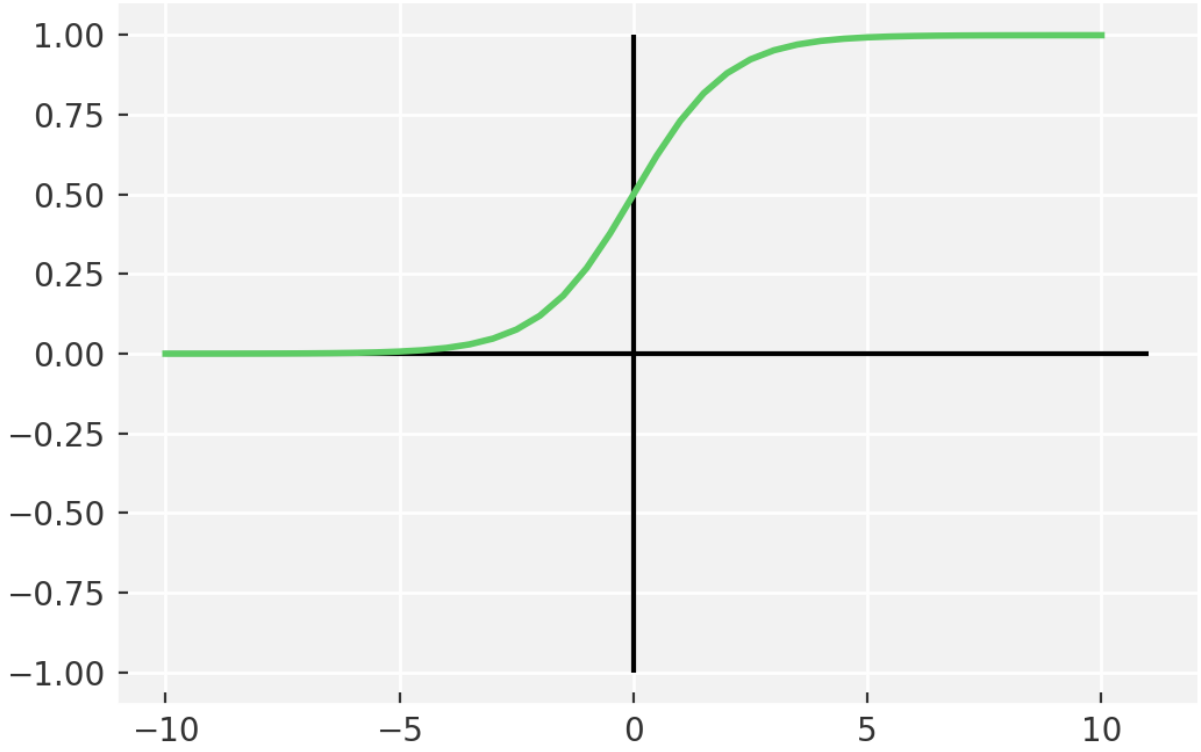
- The simplest form of activation function



Science and Technology Facilities Council

Activation function: Sigmoid

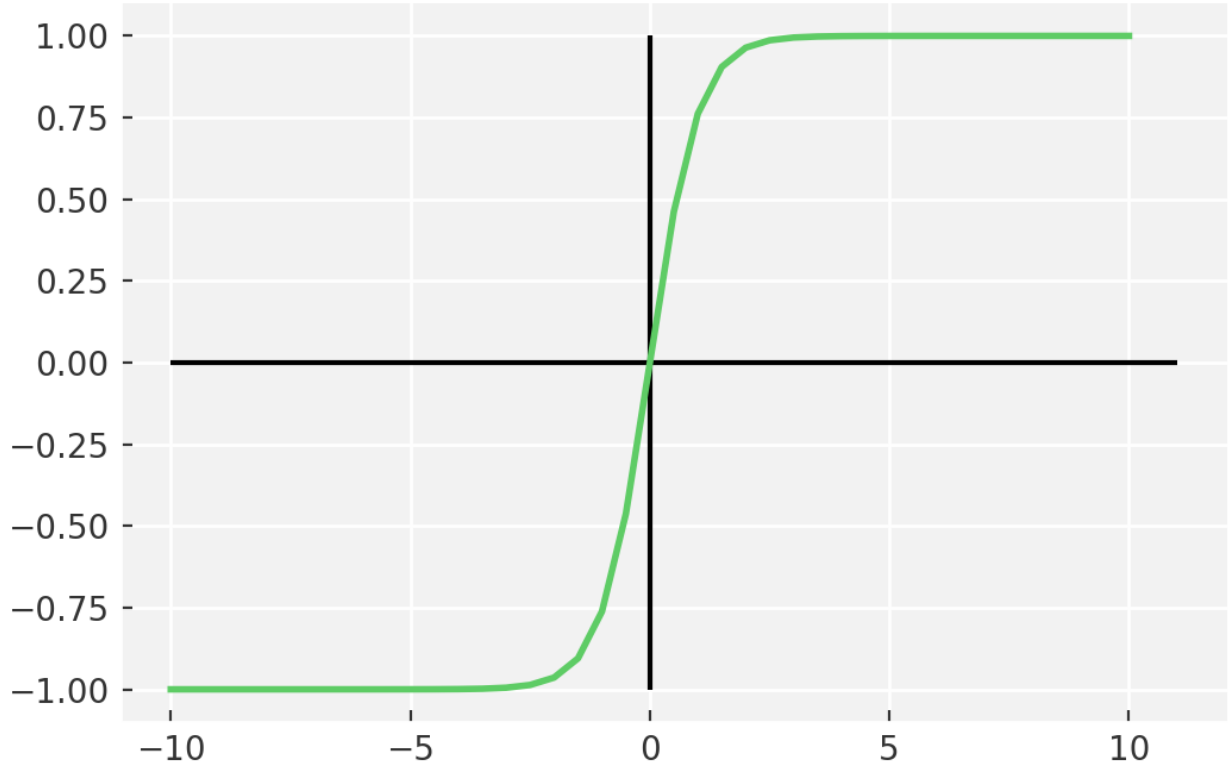
- Vanishing gradient problem
- Secondly , its output isn’t zero centered. It makes the gradient updates go too far in different directions. **0 < output < 1, and it makes optimization harder.**
- Sigmoids saturate and kill gradients.
- Sigmoids have slow convergence.



Science and Technology Facilities Council

Activation function: Tanh

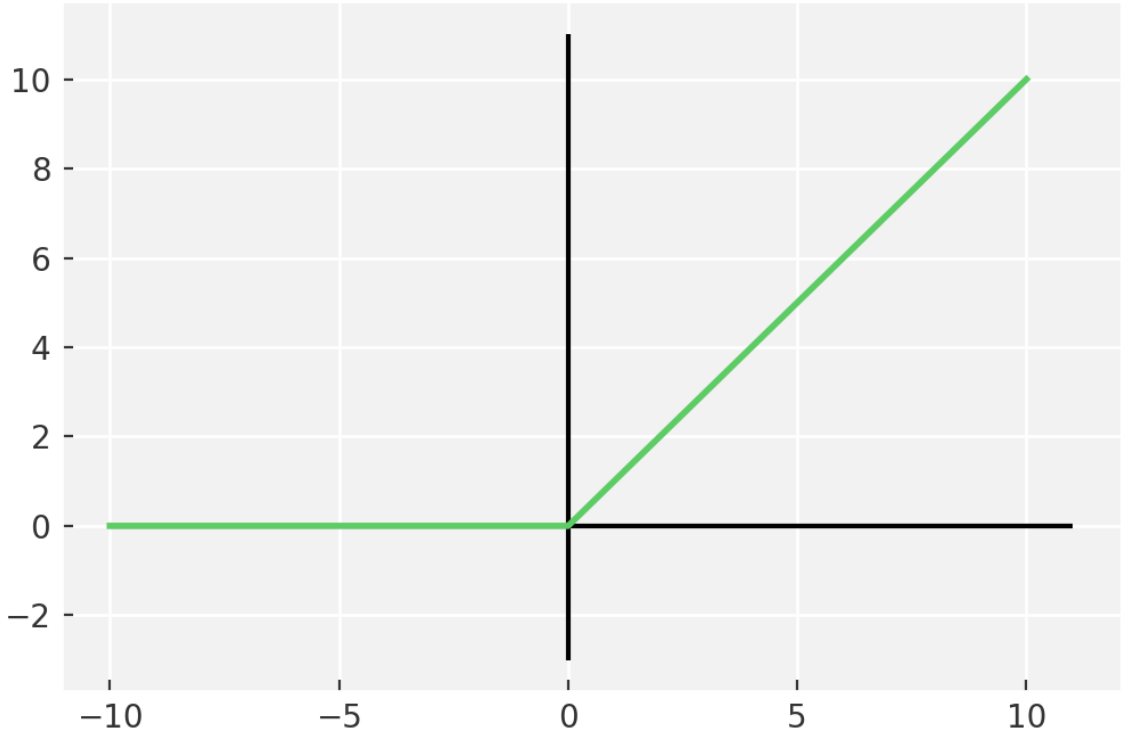
- Output is zero centered
- Usually preferred to sigmoid as it converges better
- Still it suffers from vanishing gradient problem



Science and Technology Facilities Council

Activation function: ReLU

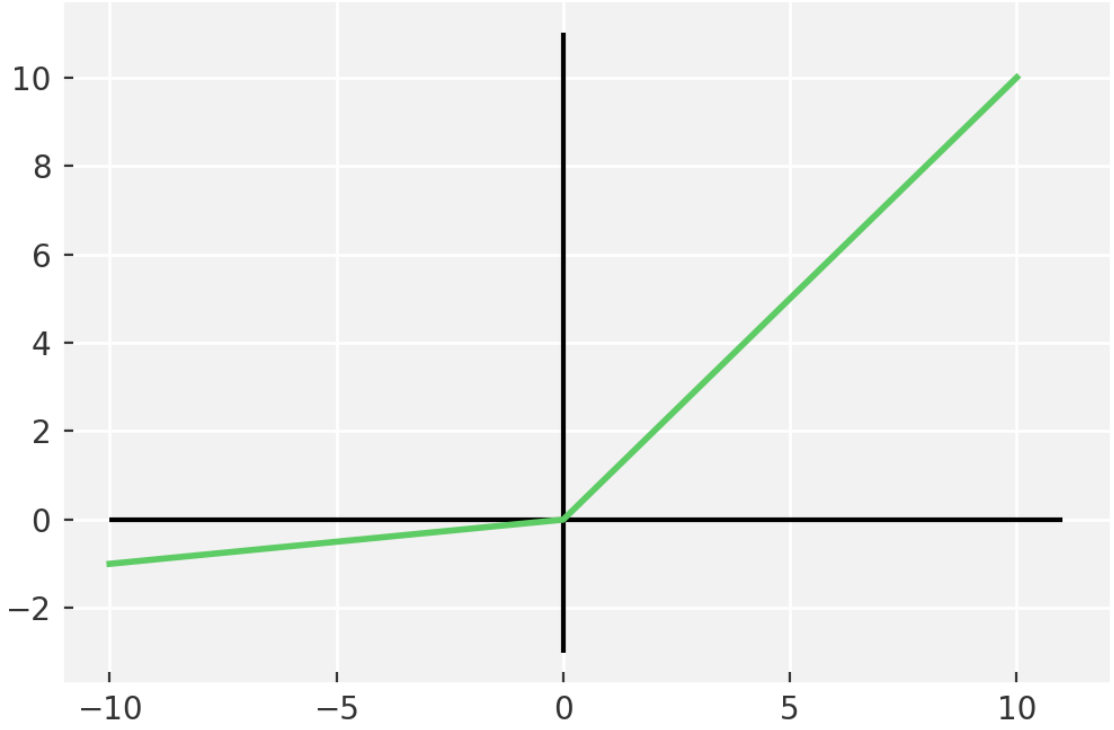
- 6 times improvement in convergence from Tanh function
- Should only be used within Hidden layers of a neural network model



Science and Technology Facilities Council

Activation function: LeakyReLU

- Some ReLu gradients can be fragile during training and can die.
- Cause a weight update which will makes it never activate on any data point again.
- ReLu could result in Dead Neurons.




```
keras.layers.LeakyReLU(alpha=0.3)
```

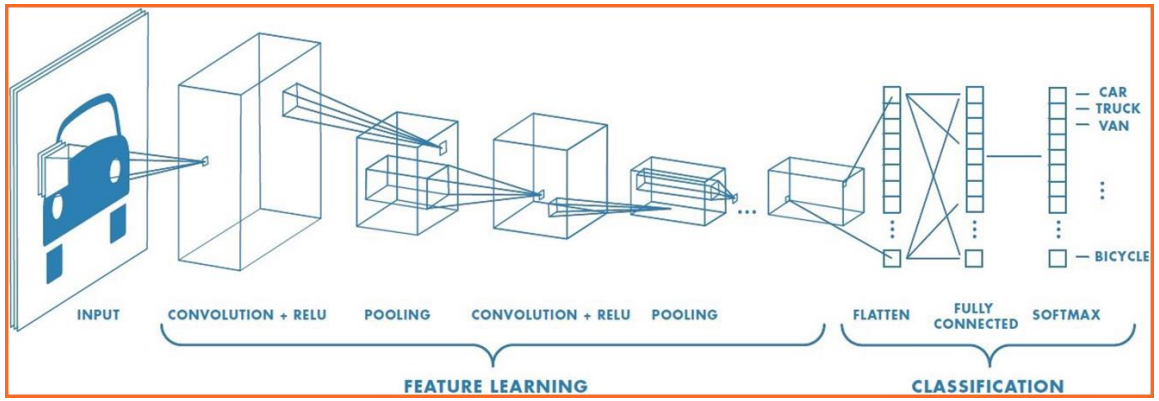
Science and Technology Facilities Council

Go To Notebook

- Its time to build your first neural network

Neural networks

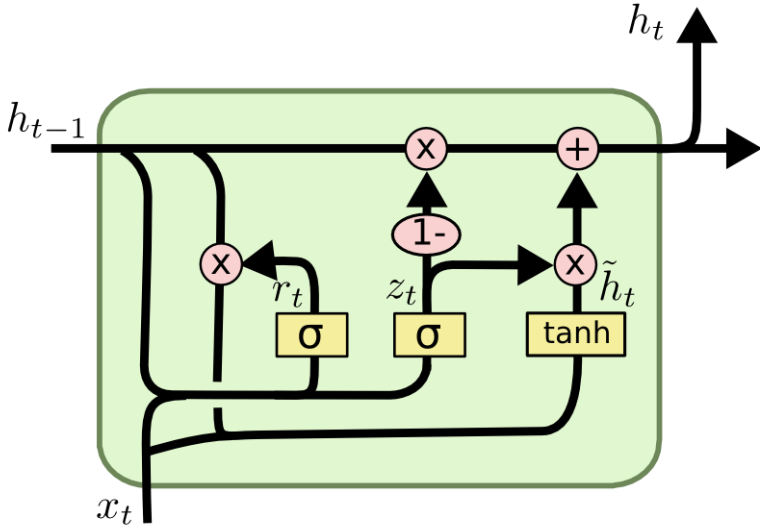
- CNNs image
- The MLP has no real concept of the spatial relations
- Also, dense connections lead to parametric explosions for many pixel images



Science and Technology Facilities Council

Neural networks

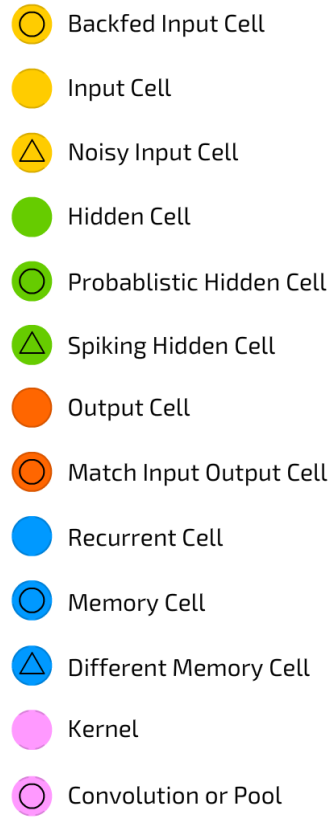
- RNNs/LSTMs time series
  - Often algorithms are desired for predicting the next event based on a series of previous events
  - Eg Pressure/temperature evolution, speech prediction …
  - In this case standard NNs are not very useful due to a lack of ‘memory’

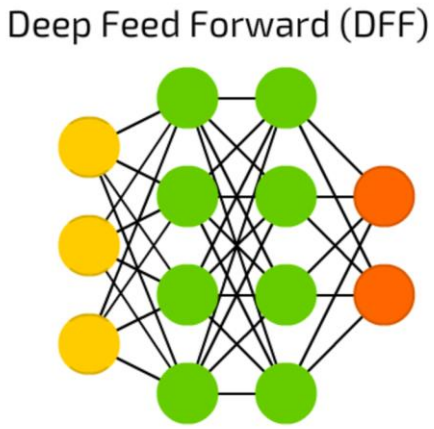


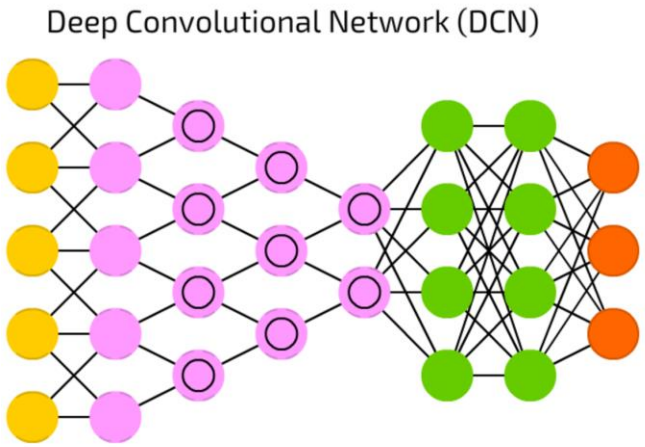
Science and Technology Facilities Council

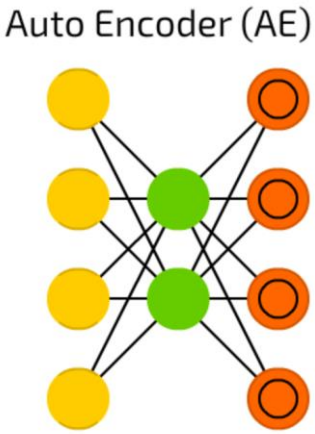
https://colah.github.io/posts/2015-08-Understanding-LSTMs/

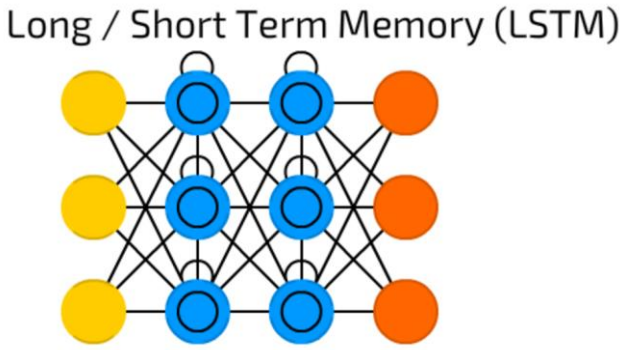
Common types of NN











Science and Technology Facilities Co

Concept checklist

- Origins of neural networks
- Types of layers – input, hidden, output
- Dense/fully-connected layers
- Activation functions
- Batch/epoch
- Dropout

Concept checklist

- Supervised/unsupervised machine learning
- Classical machine learning/deep learning
- Parameters/hyperparameters
- Features and feature engineering
- Decision trees
- Overfitting
- Evaluation/metrics
- Test/train split, cross-validation
- Bagging and boosting
- Deep learning
- Neural networks

Science and Technology Facilities Council

Thank you

Science and Technology Facilities Council

Twitter:@keeeto2000