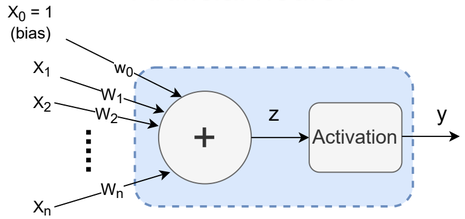
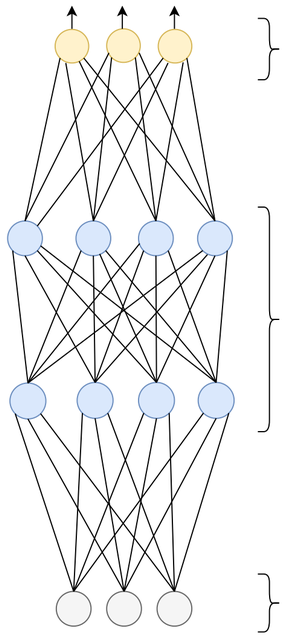
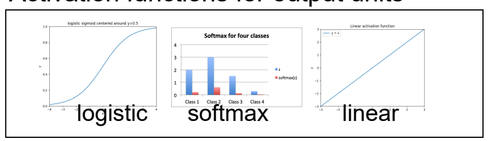
Artificial neuron



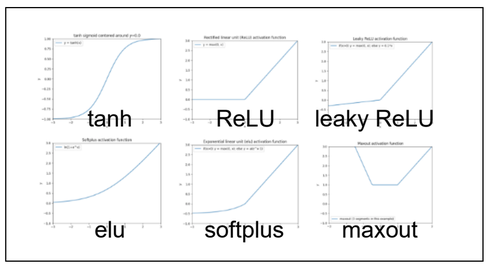
Feedforward network



Activation functions for output units



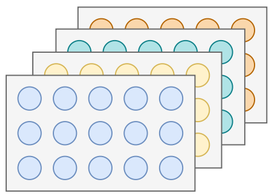
Activation functions for hidden units

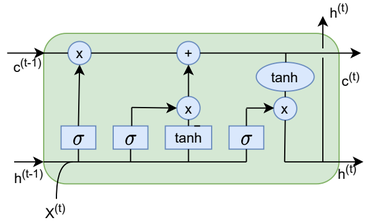


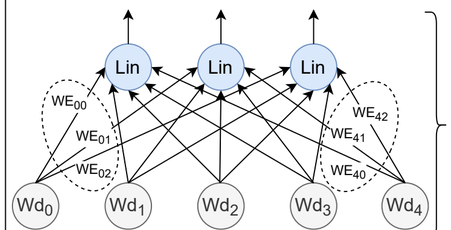
Input encodings

Standardized numerical values	One-hot encoded classes	Word embeddings from embedding layer

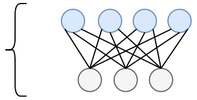
Layer types

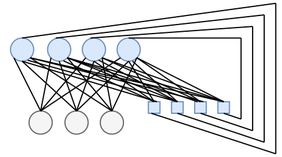


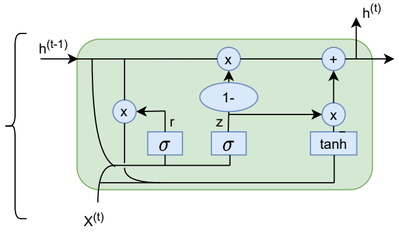


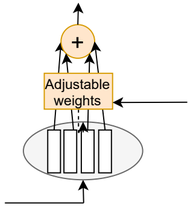


Type	Description	Example usage
Fully connected	Each neuron connects to each output in preceding layer. Also known as projection layer if no activation function is used.	Cases where specialized layers do not provide additional value
Convolutional	Sparsely connected. Employs weight sharing. Consists of multiple channels. Each channel is often arranged in two dimensions.	Image processing (2D convolution) and text processing (1D convolution)
Simple recurrent	Recurrent connections. Output from previous timestep is used as input. Weight sharing between timesteps.	Sequential data of variable length, e.g., text processing
Long short-term memory (LSTM)	Recurrent layer with more complex units. Each unit contains an internal memory cell. Gates control when to remember and forget.	Long sequences, e.g., text processing
Gated recurrent unit (GRU)	Simplified version of LSTM. No internal memory cell but still has gates to control when to remember or forget previous output value.	Long sequences, e.g., text processing
Embedding	Converts sparse one-hot encoded data into a dense representation. Implemented as lookup table.	Convert textual input data into word embeddings
Attention	Output vector is a weighted sum of multiple input vectors. The weights are dynamically chosen to attend to the most important vector.	Extract information from long text sequences or images









From *Learning Deep Learning* by Magnus Ekman (ISBN: 9780137470358, www.ldlbook.com). Copyright 2022 NVIDIA Corporation. All rights reserved.