Introduction To Deep Learning

Workshop 2

NetBrain.ml

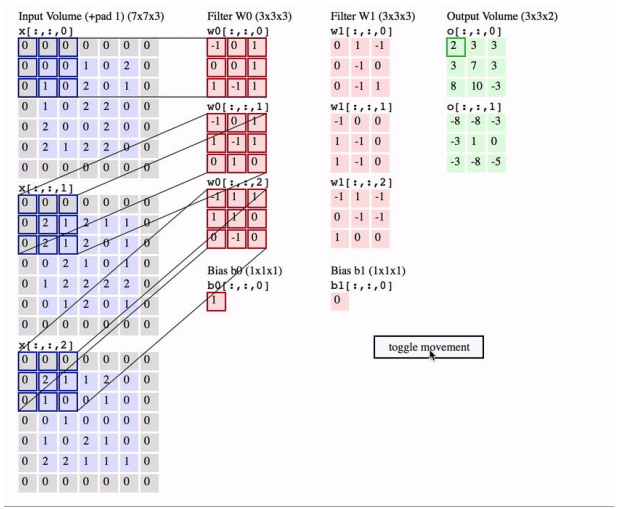
Last Edit: 11/11/19

Convolutional Networks

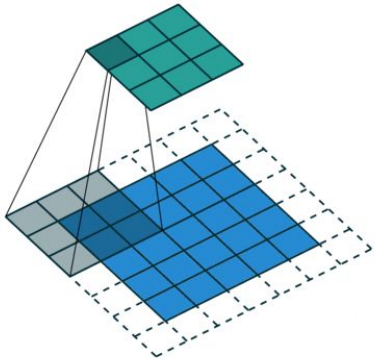
Goal**:** To handle datasets with ***dispersed but shared features***, using filter by convolution operation should be used, replacing the standard affine function.

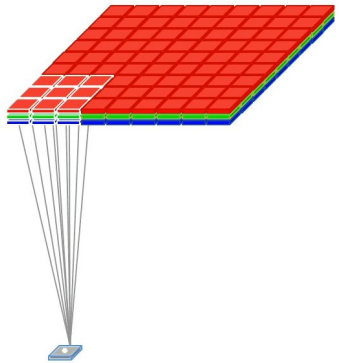
Convolution**:** Having a sliding window of parameters that does element-wise multiplication and sums the products.

Note: Convolution can be represented as a matrix multiplication



$$\begin{pmatrix} k1 & k2 & 0 & k3 & k4 & 0 & 0 & 0 & 0 \\ 0 & k1 & k2 & 0 & k3 & k4 & 0 & 0 & 0 \\ 0 & 0 & 0 & k1 & k2 & 0 & k3 & k4 & 0 \\ 0 & 0 & 0 & 0 & k1 & k2 & 0 & k3 & k4 \end{pmatrix} \cdot \begin{pmatrix} x1 \\ x2 \\ x3 \\ x4 \\ x5 \\ x6 \\ x7 \\ x8 \\ x9 \end{pmatrix}$$





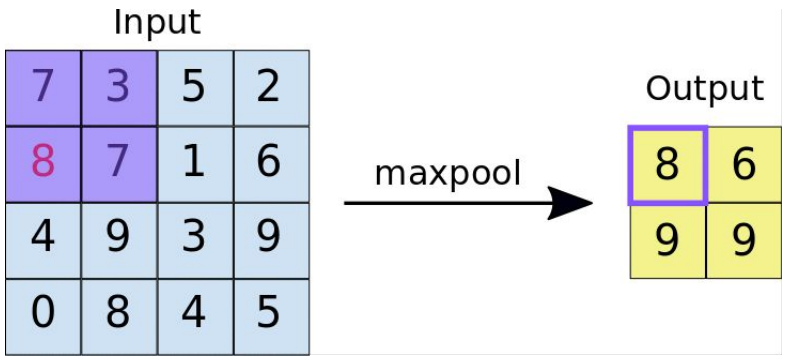
Max / Average / Adaptive Pooling Layer

Goal: ***Reduce dimensionality*** of data for computational efficiency and can improve generalization

Max Pool: Within a window, return the max value

Average Pool: Within a window, return the average

Adaptive (or Global) Pool: Take an arbitrary size as input and outputs to the appropriate size



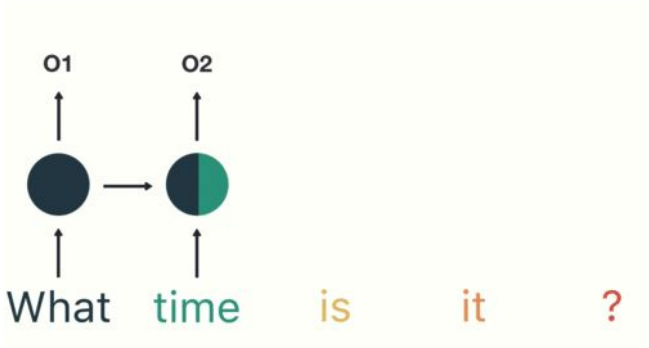
Recurrent Networks

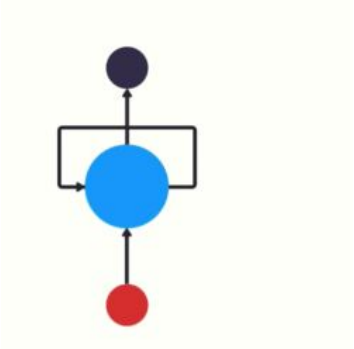
Goal: To ***handle sequential*** datasets where order matters, a feedback is used to act as memory

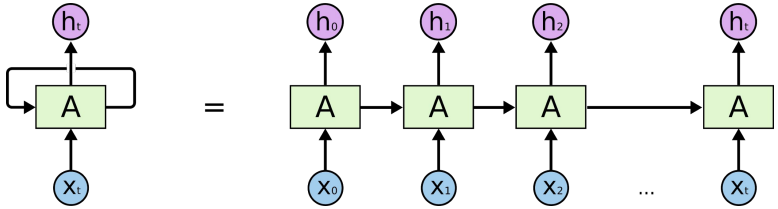
Feedback: The previous output is fed through as part of the input. The first feedback is initialized with zeros (usually).

Note: Every input is fed through the same node, using the same parameters

Learning Technique: Uses backpropagation through time, which unfolds the loop and performs backpropagation.







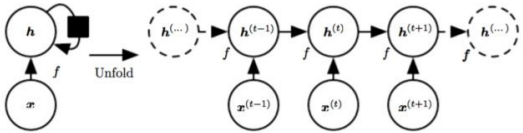


Figure 10.2: A recurrent network with no outputs. This recurrent network just processes

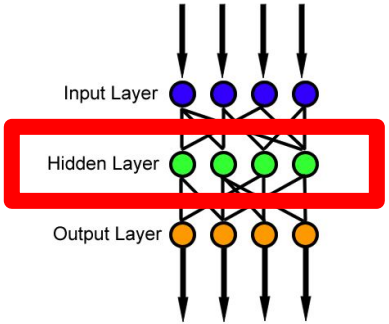
Embedding Layer

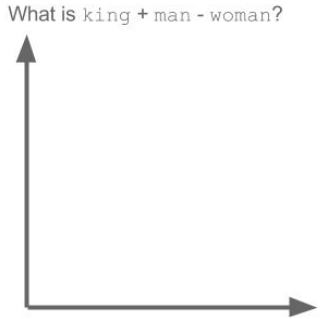
Goal: ***Develop meaning*** behind each feature/word

LDA: Learn mapping word with various topics

Topic 1		Topic 2		Topic 3	
term	**weight**	**term**	**weight**	**term**	**weight**
biology	45692	space	67019	politics	24763
university	10576	nasa	9673	washington	11982
moth	5304	earth	5674	congress	7261
caterpillar	4927	moon	3455	president	5820

Word2Vec: Uses MLP to learn words as vectors representation with contextual meaning.





FastText: Uses character n-grams, uses Word2Vec.

Ex. ‘Ride’ becomes <’rid’, ’ide’>, given n=3

GloVe: Uses word-word co-occurrence matrix, which contains word frequency in a context, and reduce it

	0	1	2	3	4	5	6	7	8
fox	-0.348680	-0.077720	0.177750	-0.094953	-0.452890	0.237790	0.209440	0.037886	0.035064
ham	-0.773320	-0.282540	0.580760	0.841480	0.258540	0.585210	-0.021890	-0.463680	0.139070
brown	-0.374120	-0.076264	0.109260	0.186620	0.029943	0.182700	-0.631980	0.133060	-0.128980
beautiful	0.171200	0.534390	-0.348540	-0.097234	0.101800	-0.170860	0.295650	-0.041816	-0.516550
jumps	-0.334840	0.215990	-0.350440	-0.260020	0.411070	0.154010	-0.386110	0.206380	0.386700
eggs	-0.417810	-0.035192	-0.126150	-0.215930	-0.669740	0.513250	-0.797090	-0.068611	0.634660
beans	-0.423290	-0.264500	0.200870	0.082187	0.066944	1.027600	-0.989140	-0.259950	0.145960
sky	0.312550	-0.303080	0.019587	-0.354940	0.100180	-0.141530	-0.514270	0.886110	-0.530540
bacon	-0.430730	-0.016025	0.484620	0.101390	-0.299200	0.761820	-0.353130	-0.325290	0.156730
breakfast	0.073378	0.227670	0.208420	-0.456790	-0.078219	0.601960	-0.024494	-0.467980	0.054627
toast	0.130740	-0.193730	0.253270	0.090102	-0.272580	-0.030571	0.096945	-0.115060	0.484000
today	-0.156570	0.594890	-0.031445	-0.077586	0.278630	-0.509210	-0.066350	-0.081890	-0.047986
blue	0.129450	0.036518	0.032298	-0.060034	0.399840	-0.103020	-0.507880	0.076630	-0.422920

ELMo: Uses bidirectional language model to learn words as vectors representation with contextual meaning. This can now look at each word in its actual context

Gatings: GRUs and LSTMs

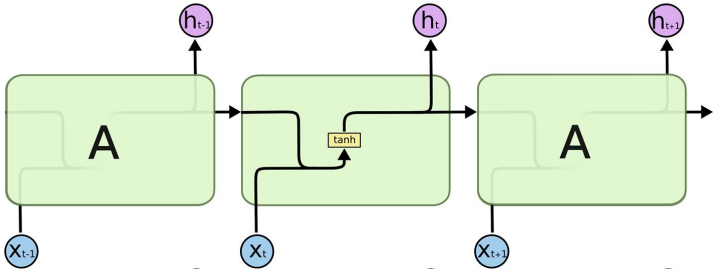
Goal: Due to ***exploding/vanishing gradients*** (EVG) in recurrent cells, gated feedback connections are introduced.

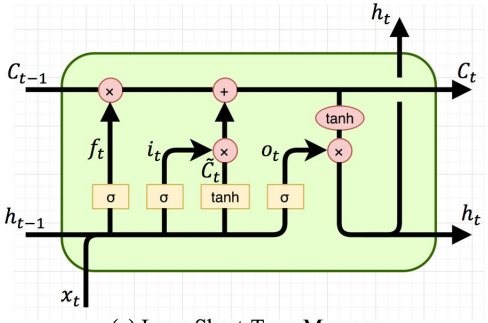
EVG: Caused by reuse of parameters and how much the parameters diverge from 1 since repeated multiplication may make the gradient infinitely small or large

Gates: Instead of affine feedback, ***nonlinearities are introduced*** to manage memory better

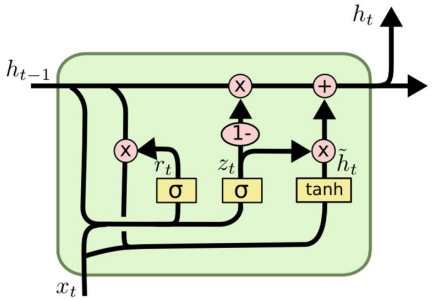
LSTMs: Forget, Input, and Output

GRUs: Reset, Update





(a) Long Short-Term Memory



(b) Gated Recurrent Unit