Deep Learning for Healthcare

Oguzhan Gencoglu - Top Data Science

GE HC Science & Technology Week - Helsinki - 22 May 2017

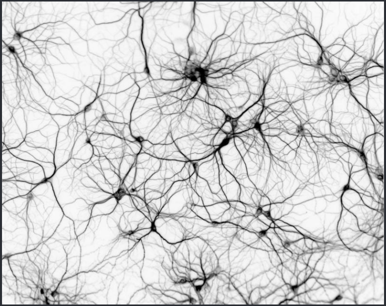
TOP data science

Outline

The Idea & Brief History

Artificial Neural Networks

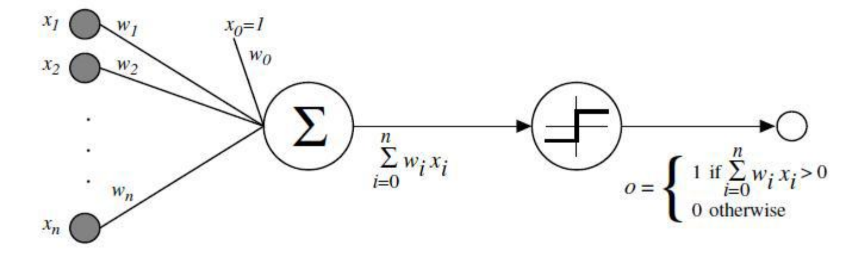
Artificial neural networks are computational models inspired by biological neural networks (in reality this is a very loose analogy).



The Idea & Brief History

Timeline

- Mcculoch-Pitts model, implementing basic AND/OR/NOT (1943).
- Rosenblatt's simplified neuron model, i.e., perceptron (1958).



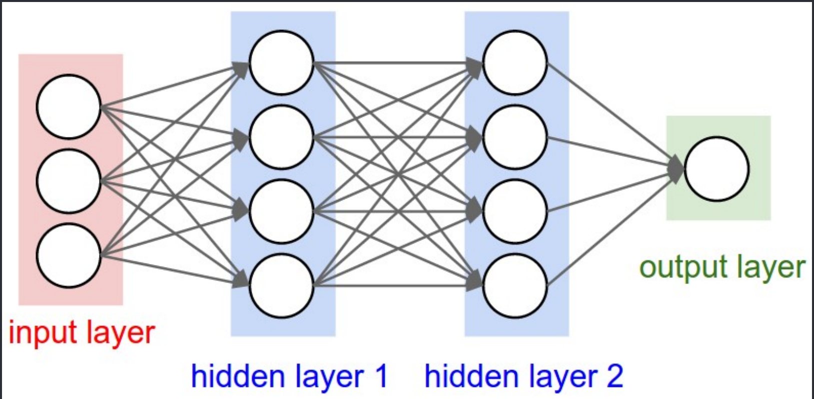
The Idea & Brief History

Timeline

- Backpropagation already derived in early 1960s but incomplete and inefficient.
- Linnainmaa formed the modern backprop in his masters thesis (1970). Did not really mention its applications to NNs.
- The famous paper by Rumelhart, Hinton, and Williams made a breakthrough.

Title 1–20	Cited by	Year
Learning internal representations by error-propagation DE Rumelhart, GE Hinton, RJ Williams Parallel Distributed Processing: Explorations in the Microstructure of ...	22855	1986

Neural Network - What exactly is it?



Neural Network - What exactly is it?

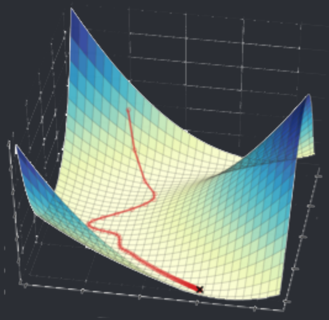
Breakdown

- Input Layer: Features
- Connections (weights): Coefficients used for sum and product calculations. Learning means modifying these coefficients (parameters).
- Hidden Layer(s): Consists of neurons with some non-linearity (activation function)
- Output Layer: Targets (single neuron if regression, *NumberOfClasses* neurons if classification)

Training Neural Nets

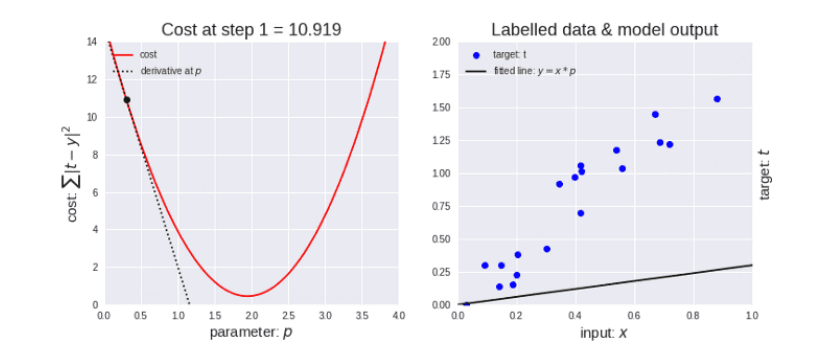
Backpropagation

- Backwards propagation of errors
- Based on gradient descent (non-convex optimization)



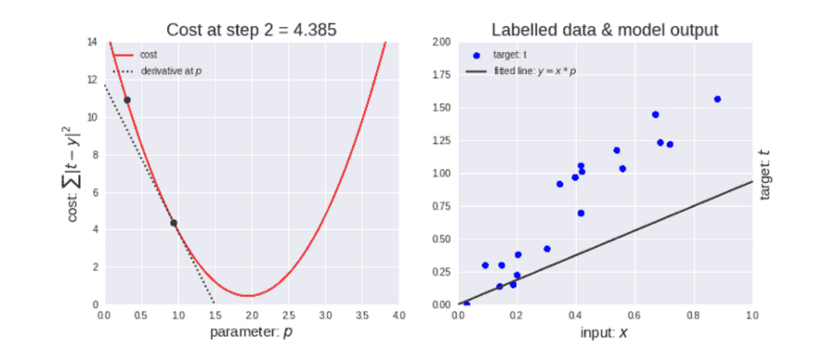
Training Neural Nets

Backpropagation



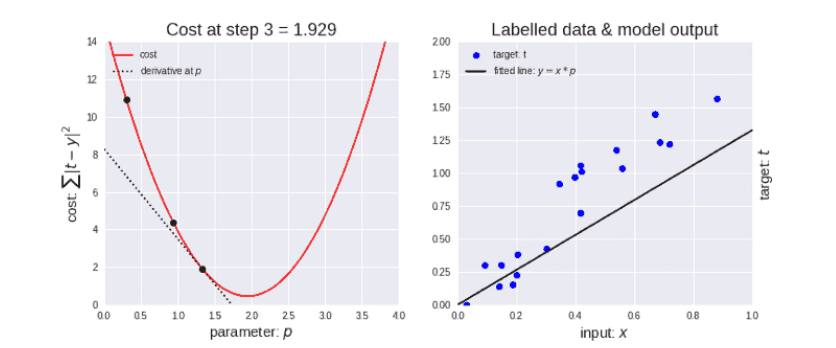
Training Neural Nets

Backpropagation



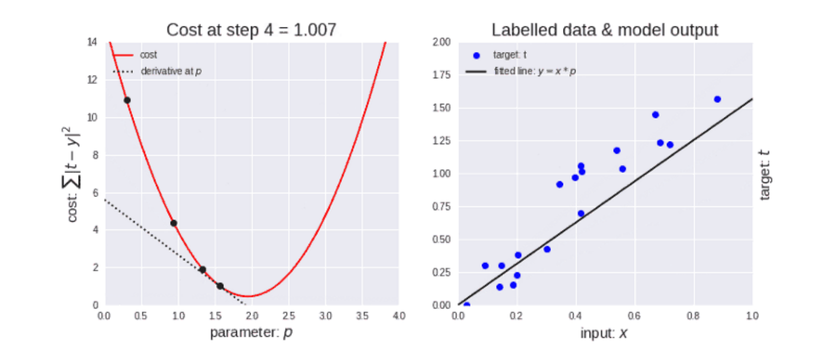
Training Neural Nets

Backpropagation



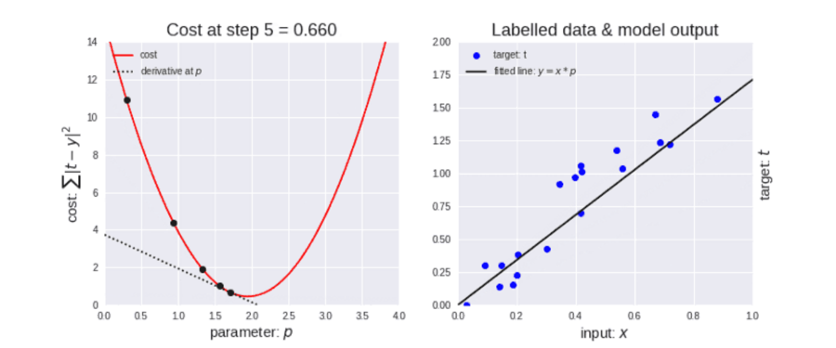
Training Neural Nets

Backpropagation



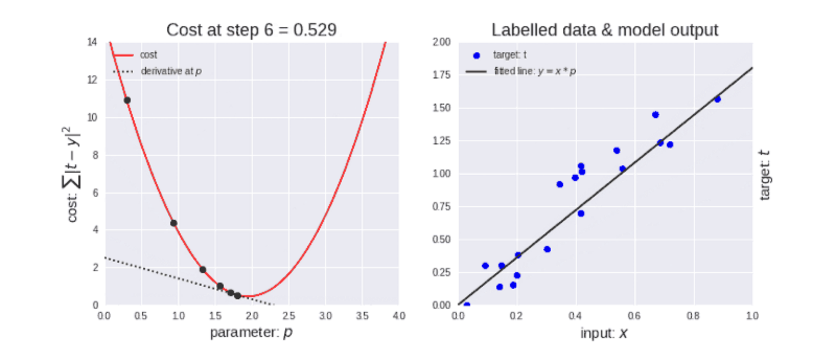
Training Neural Nets

Backpropagation



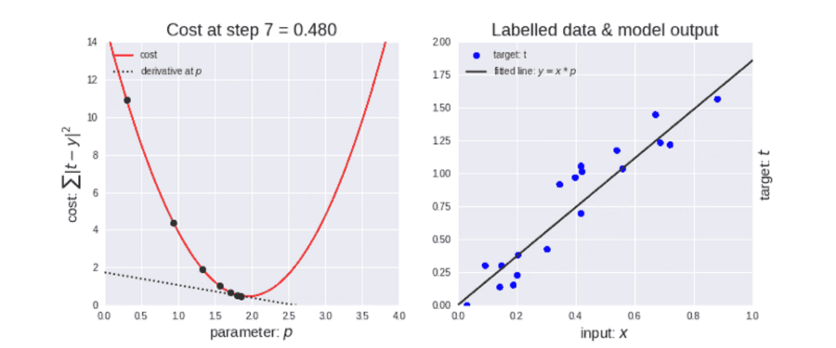
Training Neural Nets

Backpropagation



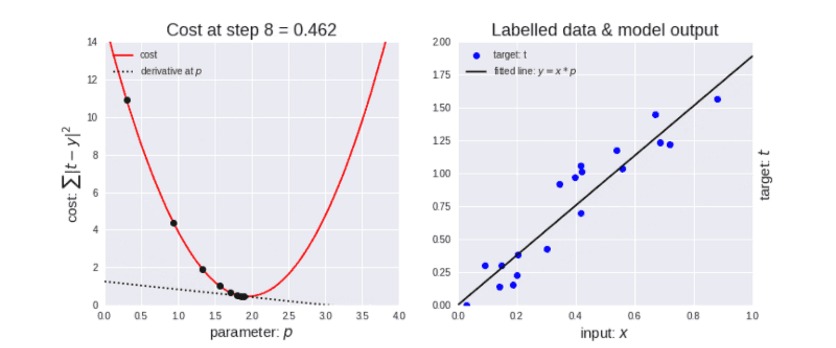
Training Neural Nets

Backpropagation



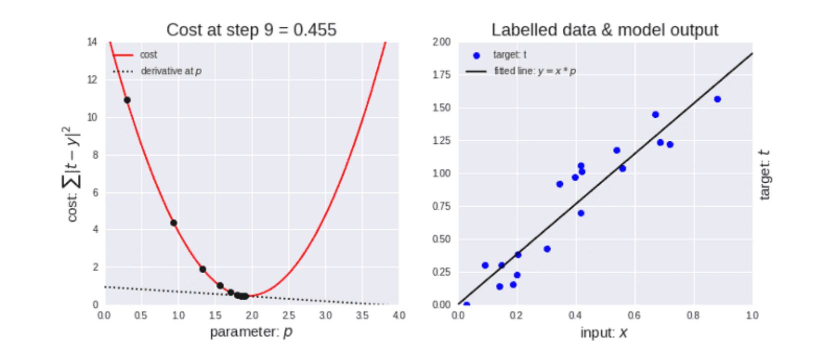
Training Neural Nets

Backpropagation



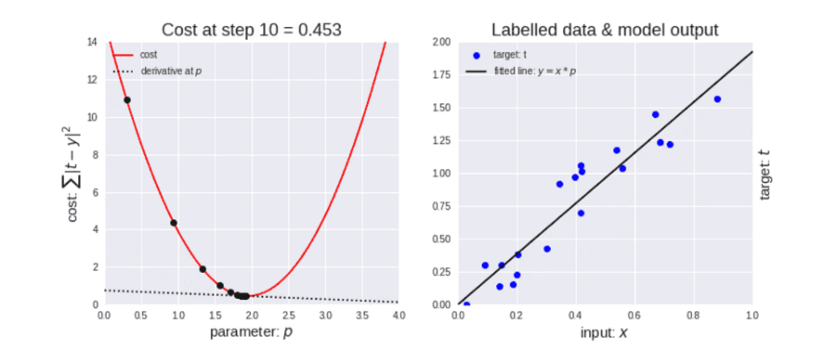
Training Neural Nets

Backpropagation



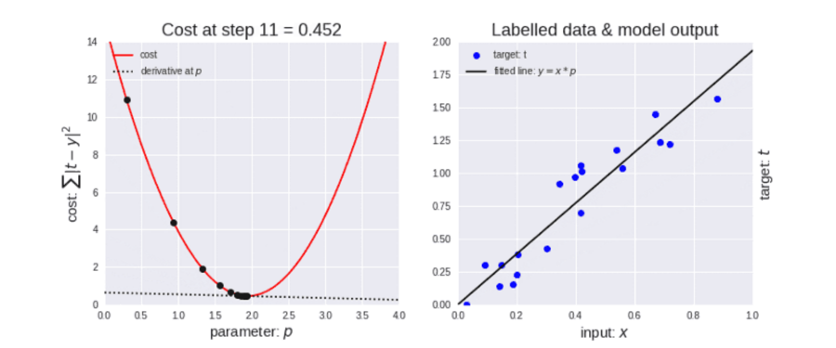
Training Neural Nets

Backpropagation



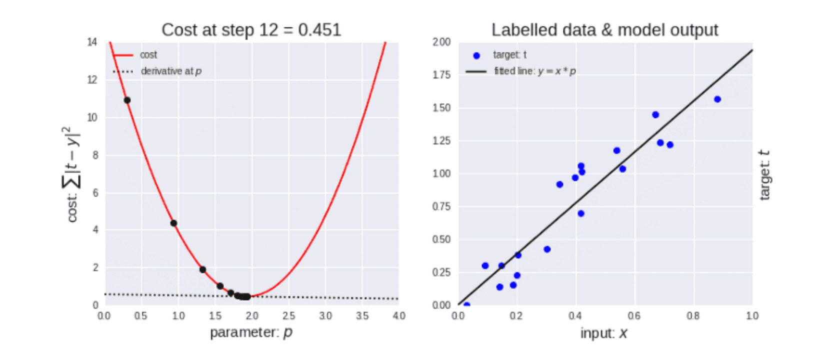
Training Neural Nets

Backpropagation



Training Neural Nets

Backpropagation



Let's Play

http://playground.tensorflow.org/

The AI Winter

- Increased hype and failure to deliver high expectations (limited knowledge on efficient training algorithms and hardware) let to The AI Winter.

The AI Winter

- Increased hype and failure to deliver high expectations (limited knowledge on efficient training algorithms and hardware) let to The AI Winter.
- The term AI droped to an almost pseudoscience status.

The AI Winter

- Increased hype and failure to deliver high expectations (limited knowledge on efficient training algorithms and hardware) let to The AI Winter.
- The term AI droped to an almost pseudoscience status.
- Funding for neural networks research went scarce.

The AI Winter

- Increased hype and failure to deliver high expectations (limited knowledge on efficient training algorithms and hardware) let to The AI Winter.
- The term AI droped to an almost pseudoscience status.
- Funding for neural networks research went scarce.
- Some important advances were made (LSTMs) but went unnoticed.

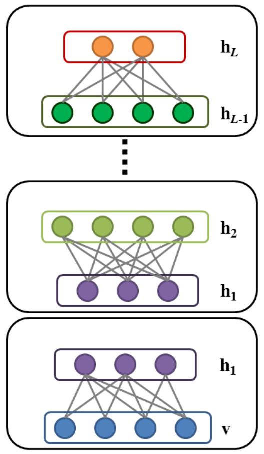
The AI Winter

- Increased hype and failure to deliver high expectations (limited knowledge on efficient training algorithms and hardware) let to The AI Winter.
- The term AI droped to an almost pseudoscience status.
- Funding for neural networks research went scarce.
- Some important advances were made (LSTMs) but went unnoticed.
- Support Vector Machines (SVM) by Cortes and Vapnik (1995) ate most of the cake.

Outline

How It All Started?

Deep Belief Networks[1]

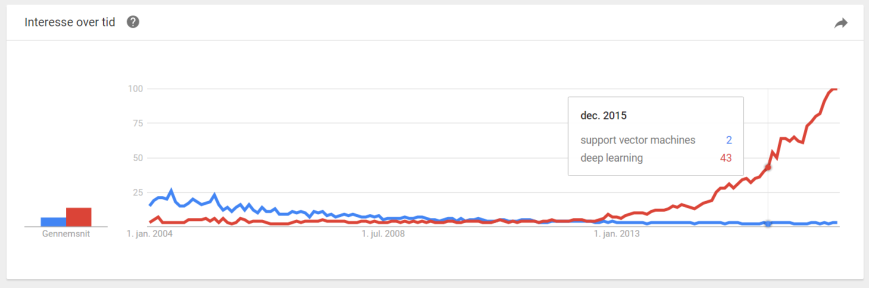


Remarks

- Greedy, layer-wise, unsupervised pretraining (contrastive divergence)
- Fine-tuning with backpropagation

[1] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural computation. 2006 Jul;18(7):1527-54.

Current Status



Current Status

Remarks

- Feature engineering is becoming obsolete, especially if you have lots of data. The main idea is representation learning.

Current Status

Remarks

- Feature engineering is becoming obsolete, especially if you have lots of data. The main idea is representation learning.
- Models have high performance and complexity, but low interpretability (black box).

Current Status

Remarks

- Feature engineering is becoming obsolete, especially if you have lots of data. The main idea is representation learning.
- Models have high performance and complexity, but low interpretability (black box).
- Computationally very expensive to train (requires GPUs). Once trained, fast to use.

Current Status

Remarks

- Feature engineering is becoming obsolete, especially if you have lots of data. The main idea is representation learning.
- Models have high performance and complexity, but low interpretability (black box).
- Computationally very expensive to train (requires GPUs). Once trained, fast to use.
- No proprietary algorithms anymore.

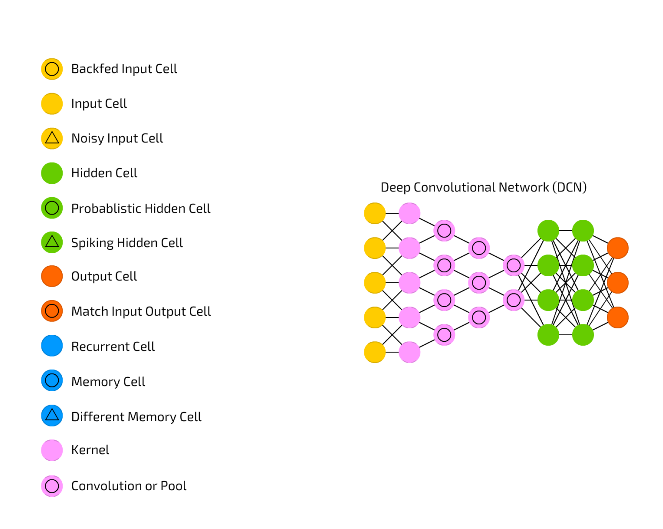
Current Status

Remarks

- Feature engineering is becoming obsolete, especially if you have lots of data. The main idea is representation learning.
- Models have high performance and complexity, but low interpretability (black box).
- Computationally very expensive to train (requires GPUs). Once trained, fast to use.
- No proprietary algorithms anymore.
- Top conferences (NIPS, ICCV, ICML) are the places to showcase.

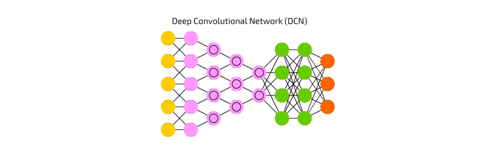
Different Architectures

Convolutional NNs



Different Architectures

Convolutional NNs[2]



Remarks

- Game changer in computer vision
- Weight sharing over a moving mask (literally a convolution in classical signal processing) thus, easier to train than feed-forward nets (less weights)
- The convolutional layers usually shrink (max-pooling trick) to help regularization and decrease number of weights

[2] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.

Deep Dive

Convolutional NNs

Convolution in classical image processing:

Operation	Kernel	Image result
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	

Deep Dive

Convolutional NNs

Convolution in classical image processing:

Operation	Kernel	Image result
Box blur (normalized)	$\frac{1}{9}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}$	
Gaussian blur 3 × 3 (approximation)	$\frac{1}{16}\begin{bmatrix}1 & 2 & 1\\ 2 & 4 & 2\\ 1 & 2 & 1\end{bmatrix}$	
Gaussian blur 5 × 5 (approximation)	$\frac{1}{256}\begin{bmatrix}1 & 4 & 6 & 4 & 1\\ 4 & 16 & 24 & 16 & 4\\ 6 & 24 & 36 & 24 & 6\\ 4 & 16 & 24 & 16 & 4\\ 1 & 4 & 6 & 4 & 1\end{bmatrix}$	
Unsharp masking 5 × 5 Based on Gaussian blur with amount as 1 and threshold as 0 (with no image mask)	$\frac{-1}{256}\begin{bmatrix}1 & 4 & 6 & 4 & 1\\ 4 & 16 & 24 & 16 & 4\\ 6 & 24 & -476 & 24 & 6\\ 4 & 16 & 24 & 16 & 4\\ 1 & 4 & 6 & 4 & 1\end{bmatrix}$	

Deep Dive

Convolutional NNs

CNN eventually learns several filters (kernels):

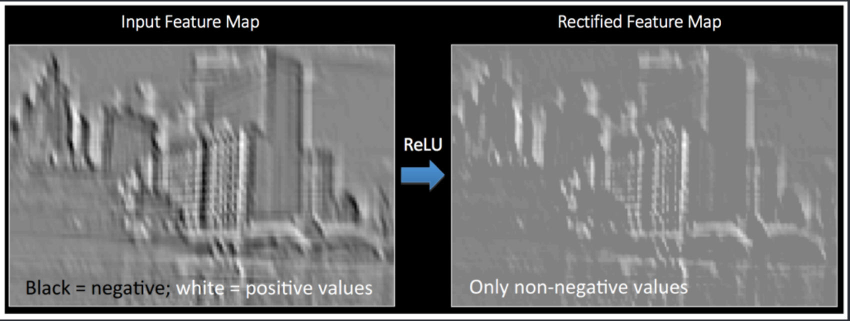
[3]

[3] If your pdf viewer does not support animation, please view it here:
https://ujwlkarn.files.wordpress.com/2016/08/giphy.gif

Deep Dive

Convolutional NNs

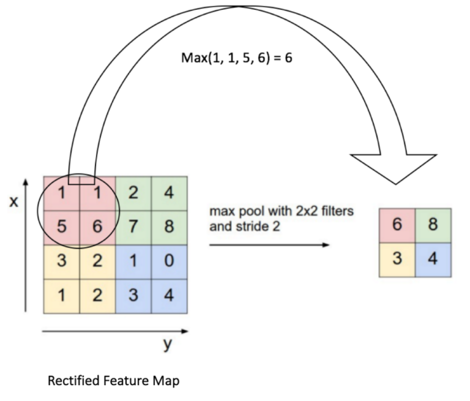
ReLU as activation function (nonlinearity):



Deep Dive

Convolutional NNs

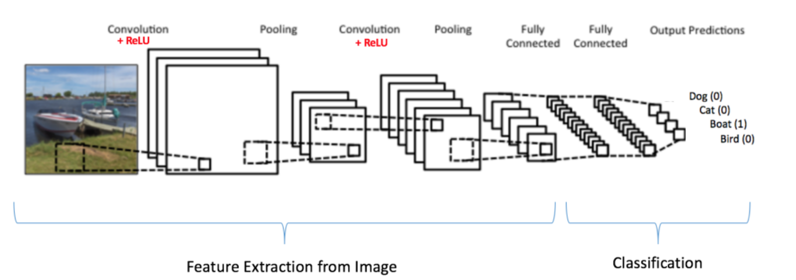
Pooling (*Max* for example):



Deep Dive

Convolutional NNs

Bringing it all together:



Game Changers

GPUs

- Ciresan and Schmidhuber used GPUs for NN training in 2010.

Game Changers

GPUs

- Ciresan and Schmidhuber used GPUs for NN training in 2010.
- NN training involves huge amounts of matrix operations.

Game Changers

GPUs

- Ciresan and Schmidhuber used GPUs for NN training in 2010.
- NN training involves huge amounts of matrix operations.
- A CPU core is much powerful than a GPU core but GPUs can have thousands of cores (CPUs: 8, 12 or in similar range) to do the work in parallel.

Game Changers

GPUs

- Ciresan and Schmidhuber used GPUs for NN training in 2010.
- NN training involves huge amounts of matrix operations.
- A CPU core is much powerful than a GPU core but GPUs can have thousands of cores (CPUs: 8, 12 or in similar range) to do the work in parallel.
- Nvidia is dominating the market.

Game Changers

AlexNet (2012)

- ILSVRC-2012 winner: 15.3% top-5 error vs. the best so far, 26.2%

Game Changers

AlexNet (2012)

- ILSVRC-2012 winner: 15.3% top-5 error vs. the best so far, 26.2%
- Left-Right flip

Game Changers

AlexNet (2012)

- ILSVRC-2012 winner: 15.3% top-5 error vs. the best so far, 26.2%
- Left-Right flip
- 8 layers: 5 conv + 3 fully connected

Game Changers

AlexNet (2012)

- ILSVRC-2012 winner: 15.3% top-5 error vs. the best so far, 26.2%
- Left-Right flip
- 8 layers: 5 conv + 3 fully connected
- ReLU, dropout, overlapping max-pooling

Game Changers

AlexNet (2012)

- ILSVRC-2012 winner: 15.3% top-5 error vs. the best so far, 26.2%
- Left-Right flip
- 8 layers: 5 conv + 3 fully connected
- ReLU, dropout, overlapping max-pooling
- 60M parameters

Game Changers

AlexNet (2012)

- ILSVRC-2012 winner: 15.3% top-5 error vs. the best so far, 26.2%
- Left-Right flip
- 8 layers: 5 conv + 3 fully connected
- ReLU, dropout, overlapping max-pooling
- 60M parameters
- Trained on NVIDIA GTX 580 GPUs

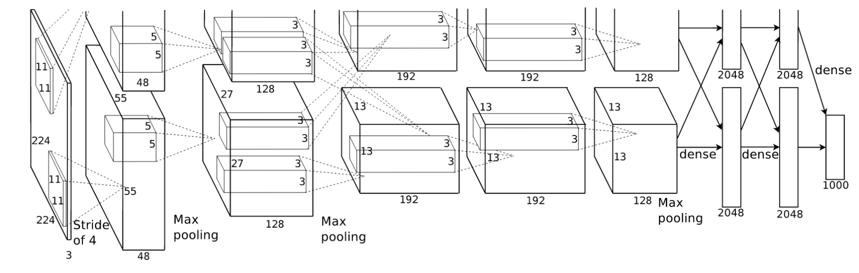
Game Changers

AlexNet (2012)

- ILSVRC-2012 winner: 15.3% top-5 error vs. the best so far, 26.2%
- Left-Right flip
- 8 layers: 5 conv + 3 fully connected
- ReLU, dropout, overlapping max-pooling
- 60M parameters
- Trained on NVIDIA GTX 580 GPUs
- Made CNNs mainstream

Game Changers

AlexNet (2012)



Game Changers

Others

- ILSVRC-2015 - Deep Residual Network

[4] Glorot (Xavier) Initialization Glorot, Xavier, and Yoshua Bengio. "Understanding the difficulty of training deep feedforward neural networks." Aistats. Vol. 9. 2010.

[5] He, Kaiming, et al. "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification." Proceedings of the IEEE international conference on computer vision. 2015.

Game Changers

Others

- ILSVRC-2015 - Deep Residual Network
- Dropout Regularization

[4] Glorot (Xavier) Initialization Glorot, Xavier, and Yoshua Bengio. "Understanding the difficulty of training deep feedforward neural networks." Aistats. Vol. 9. 2010.

[5] He, Kaiming, et al. "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification." Proceedings of the IEEE international conference on computer vision. 2015.

Game Changers

Others

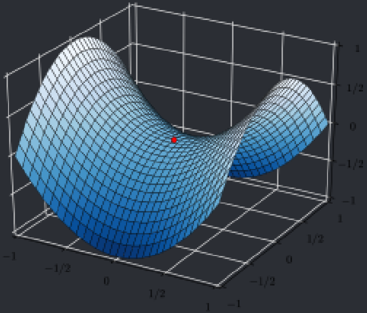
- ILSVRC-2015 - Deep Residual Network
- Dropout Regularization
- Glorot[4] & He[5] Initialization

[4] Glorot (Xavier) Initialization Glorot, Xavier, and Yoshua Bengio. "Understanding the difficulty of training deep feedforward neural networks." Aistats. Vol. 9. 2010.

[5] He, Kaiming, et al. "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification." Proceedings of the IEEE international conference on computer vision. 2015.

Issue of Local Minima in High Dimensions

Bengio's team showed that in high dimensions (larger networks) the main obstacle to training is saddle points instead of local minima[6].



[6] Dauphin, Yann N., et al. "Identifying and attacking the saddle point problem in high-dimensional non-convex optimization." Advances in neural information processing systems. 2014.

Issue of Local Minima in High Dimensions

The ratio of the number of saddle points to local minima increases exponentially with the dimensionality[7].

[7] Dauphin, Yann N., et al. "Identifying and attacking the saddle point problem in high-dimensional non-convex optimization." Advances in neural information processing systems. 2014.

Issue of Local Minima in High Dimensions

The ratio of the number of saddle points to local minima increases exponentially with the dimensionality[7].

So what?

If you wait enough (training gets very slow around saddle points), you will move on.

[7] Dauphin, Yann N., et al. "Identifying and attacking the saddle point problem in high-dimensional non-convex optimization." Advances in neural information processing systems. 2014.

Issue of Local Minima in High Dimensions

The probability of a point being a local minimum, but not a global minimum in a given dimension, increases as the loss function gets closer to the global minimum[8].

[8] Dauphin, Yann N., et al. "Identifying and attacking the saddle point problem in high-dimensional non-convex optimization." Advances in neural information processing systems. 2014.

Issue of Local Minima in High Dimensions

The probability of a point being a local minimum, but not a global minimum in a given dimension, increases as the loss function gets closer to the global minimum[8].

So what?

If you actually end up in a local minima, it is almost as good as a global minima.

[8] Dauphin, Yann N., et al. "Identifying and attacking the saddle point problem in high-dimensional non-convex optimization." Advances in neural information processing systems. 2014.

Quick Start

- Classification: softmax for last layer and cross-entropy for cost function
- Regression: Sigmoid or tanh for last layer and MSE for cost function

Quick Start

- Classification: softmax for last layer and cross-entropy for cost function
- Regression: Sigmoid or tanh for last layer and MSE for cost function
- Use ReLU as activation for hidden layers

Quick Start

- Classification: softmax for last layer and cross-entropy for cost function
- Regression: Sigmoid or tanh for last layer and MSE for cost function
- Use ReLU as activation for hidden layers
- Use Dropout for regularization

Quick Start

- Classification: softmax for last layer and cross-entropy for cost function
- Regression: Sigmoid or tanh for last layer and MSE for cost function
- Use ReLU as activation for hidden layers
- Use Dropout for regularization
- Use Adam/AdaGrad or momentum instead of plain gradient descent

Outline

Lung Cancer Detection

DATA SCIENCE BOWL.

Passion. Curiosity. Purpose.

Presented by

Booz | Allen | Hamilton & kaggle

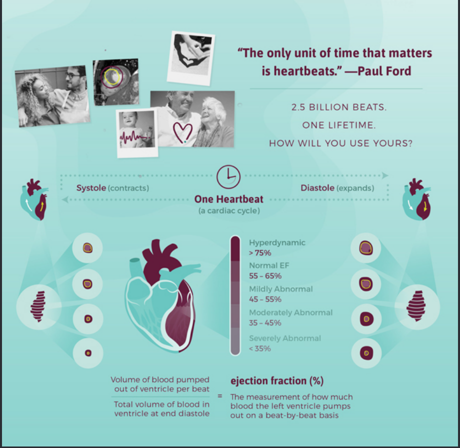
`https://www.kaggle.com/c/data-science-bowl-2017`

Prostate Cancer Detection - TDS & HUS

Voice Test for Concussion

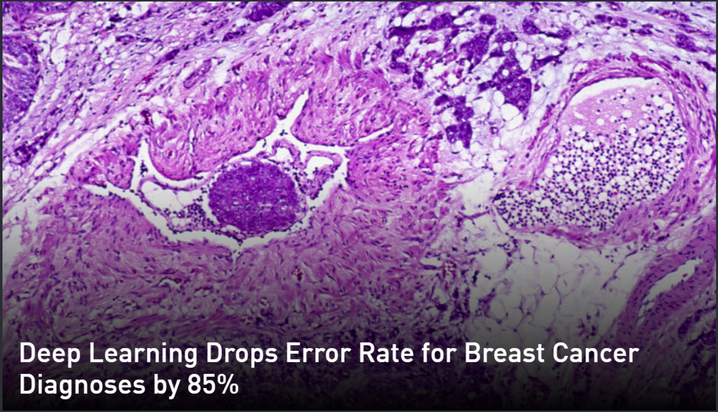
https://blogs.nvidia.com/blog/2017/03/15/ai-voice-test-for-concussion/

Diagnosing Heart Diseases



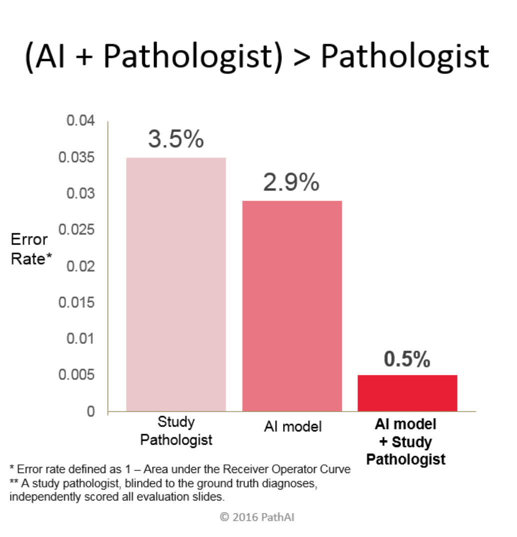
https://www.kaggle.com/c/second-annual-data-science-bowl

Breast Cancer Diagnosis

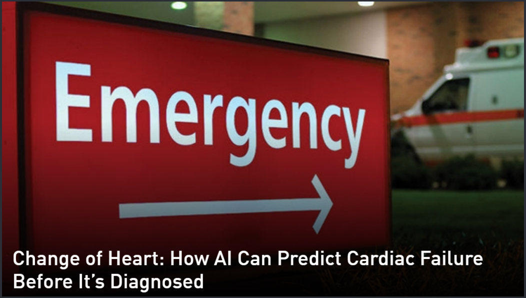


https://blogs.nvidia.com/blog/2016/09/19/deep-learning-breast-cancer-diagnosis/

Breast Cancer Diagnosis



Cardiac Failure Detection



Unlike traditional machine learning, **deep learning** does not require a human expert to define every factor the computer should evaluate in the data — a time-intensive process. In earlier research, Sun[9] said he and others spent a couple of years working with experts to build machine learning models. In three months, we were able to outperform what we had done.

[9] http://www.cc.gatech.edu/people/jimeng-sun

Diabetic Retinopathy Detection

https://www.kaggle.com/c/diabetic-retinopathy-detection

There is No Free Lunch

tekniikka &talous | TEKNIIKAN HISTORIA | metalli-tekniikka

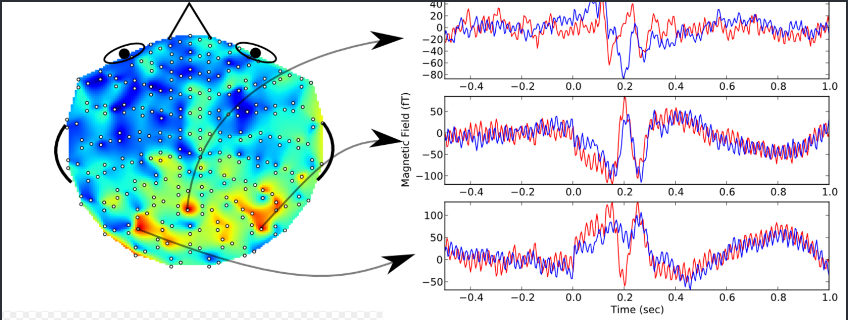
Tilaa lehti | Tilaa autokirje

ETUSIVU | UUTISET | PUHEENVUOROT | KUMPPANIBLOGIT | OTA YHTEYTTÄ

TIETOJENKÄSITTELY | Tapio Ikkala | 23.8. klo 11:34

Koneoppimiskilpailun finalisti Tampereelta - Yhden miehen tiimi haastaa IBM:n Watsonin

There is No Free Lunch



Decoding the Human Brain - 2nd Place

What to Expect in the Future?

Sooner or Later

- Automation

What to Expect in the Future?

Sooner or Later

- Automation
- More Automation

What to Expect in the Future?

Sooner or Later

- Automation
- More Automation
- Generative Models

What to Expect in the Future?

Sooner or Later

- Automation
- More Automation
- Generative Models
- End-to-end training

What to Expect in the Future?

Sooner or Later

- Automation
- More Automation
- Generative Models
- End-to-end training
- Transfer learning

What to Expect in the Future?

Sooner or Later

- Automation
- More Automation
- Generative Models
- End-to-end training
- Transfer learning
- Computer science & neuroscience collaboration

What to Expect in the Future?

Sooner or Later

- Automation
- More Automation
- Generative Models
- End-to-end training
- Transfer learning
- Computer science & neuroscience collaboration
- High-performance unsupervised learning algorithms

What to Expect in the Future?

Sooner or Later

- Automation
- More Automation
- Generative Models
- End-to-end training
- Transfer learning
- Computer science & neuroscience collaboration
- High-performance unsupervised learning algorithms
- More Automation

Popular Deep Learning Libraries

- TensorFlow
- keras
- caffe / caffe2
- PyTorch
- keras
- lasagne

Interesting Initiatives

DL, Biology and Neuroscience[10]

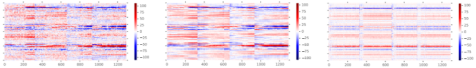
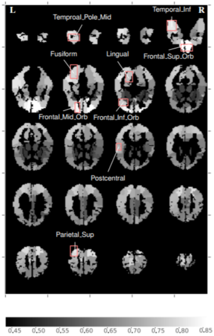


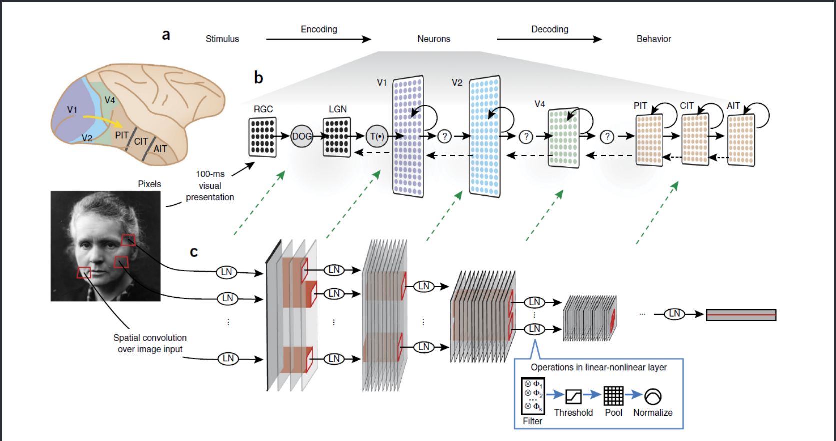
Figure 6: The real brain activity and the reconstructed brain activities over the subject_1. Four sections of the experiment are concatenated to show the complete time line of the story reading process. The pictures are constructed from (a) true brain data, (b) LSTM memory, and (c) LSTM hidden state.



[10] Qian P, Qiu X, Huang X. Bridging LSTM Architecture and the Neural Dynamics during Reading. arXiv preprint arXiv:1604.06635. 2016 Apr 22.

Interesting Initiatives

DL, Biology and Neuroscience[11]



[11] Yamins, Daniel LK, and James J. DiCarlo. "Using goal-driven deep learning models to understand sensory cortex." Nature neuroscience 19.3 (2016): 356-365.


THANK YOU