cSplash 2016

Neural Nets: Smarter than a 5th grader?

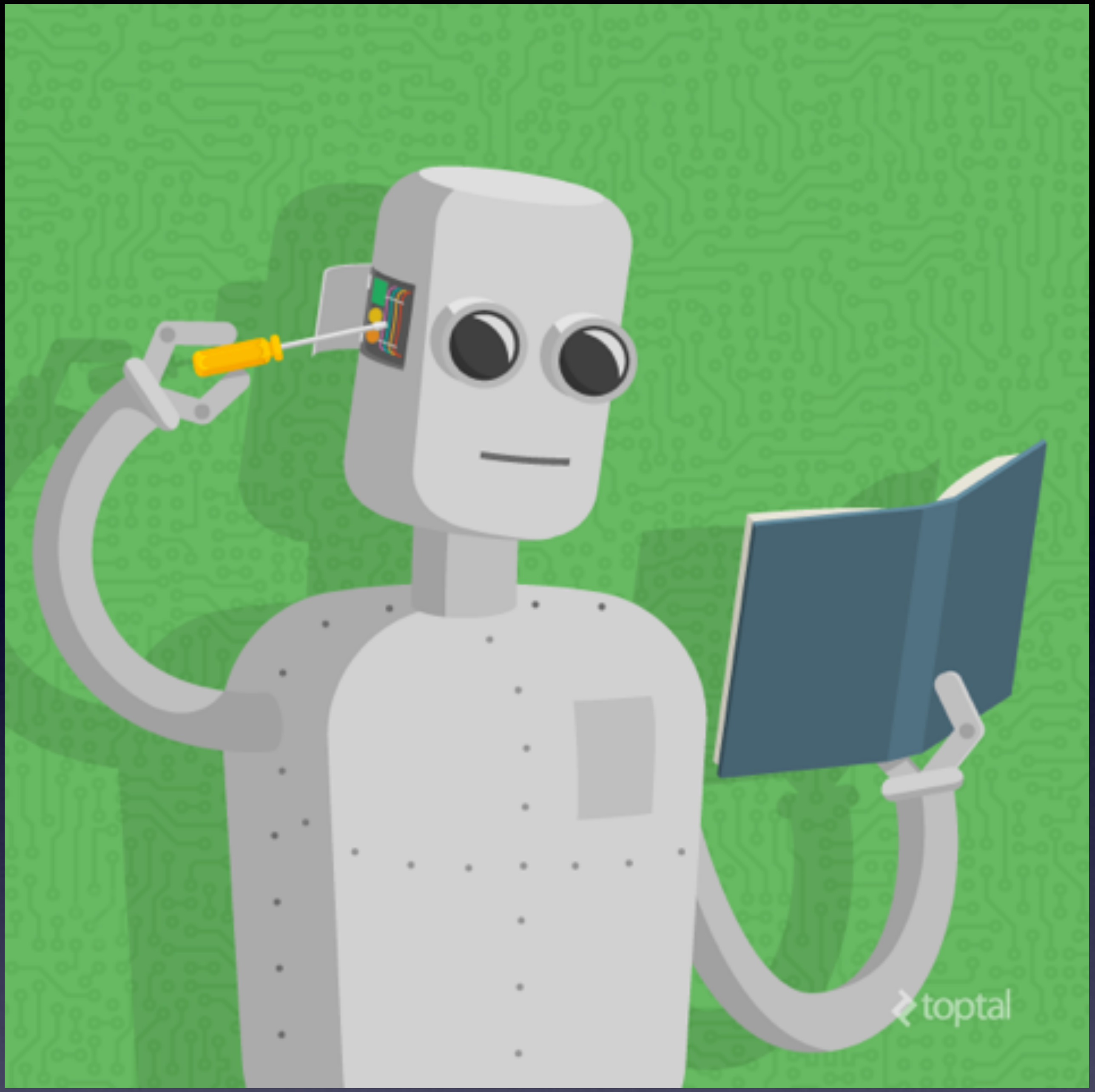
Shivam Verma (Applied Math @ Courant Institute)

@shivam13verma

The Matrix

Resources:

1. https://en.wikibooks.org/wiki/High_School_Mathematics_Extensions/Matrices
2. http://mathinsight.org/matrix_introduction
3. https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf



What is machine learning?

Image source: https://www.toptal.com/machine-learning/machine-learning-theory-an-introductory-primer

What is machine learning?

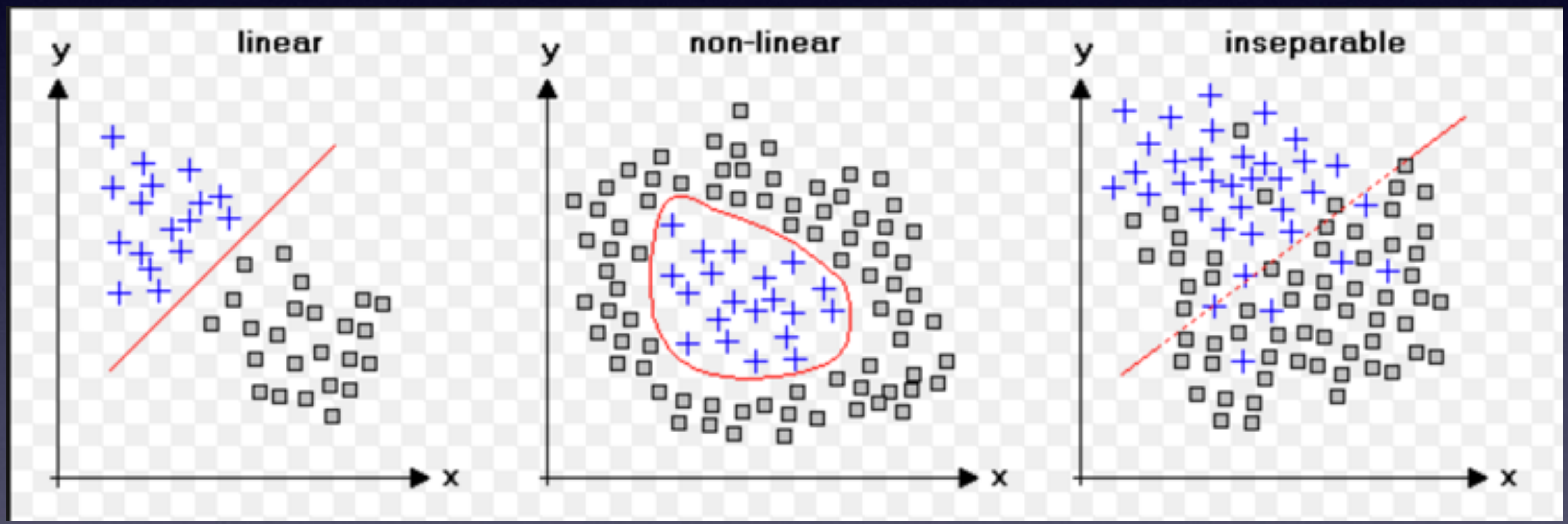
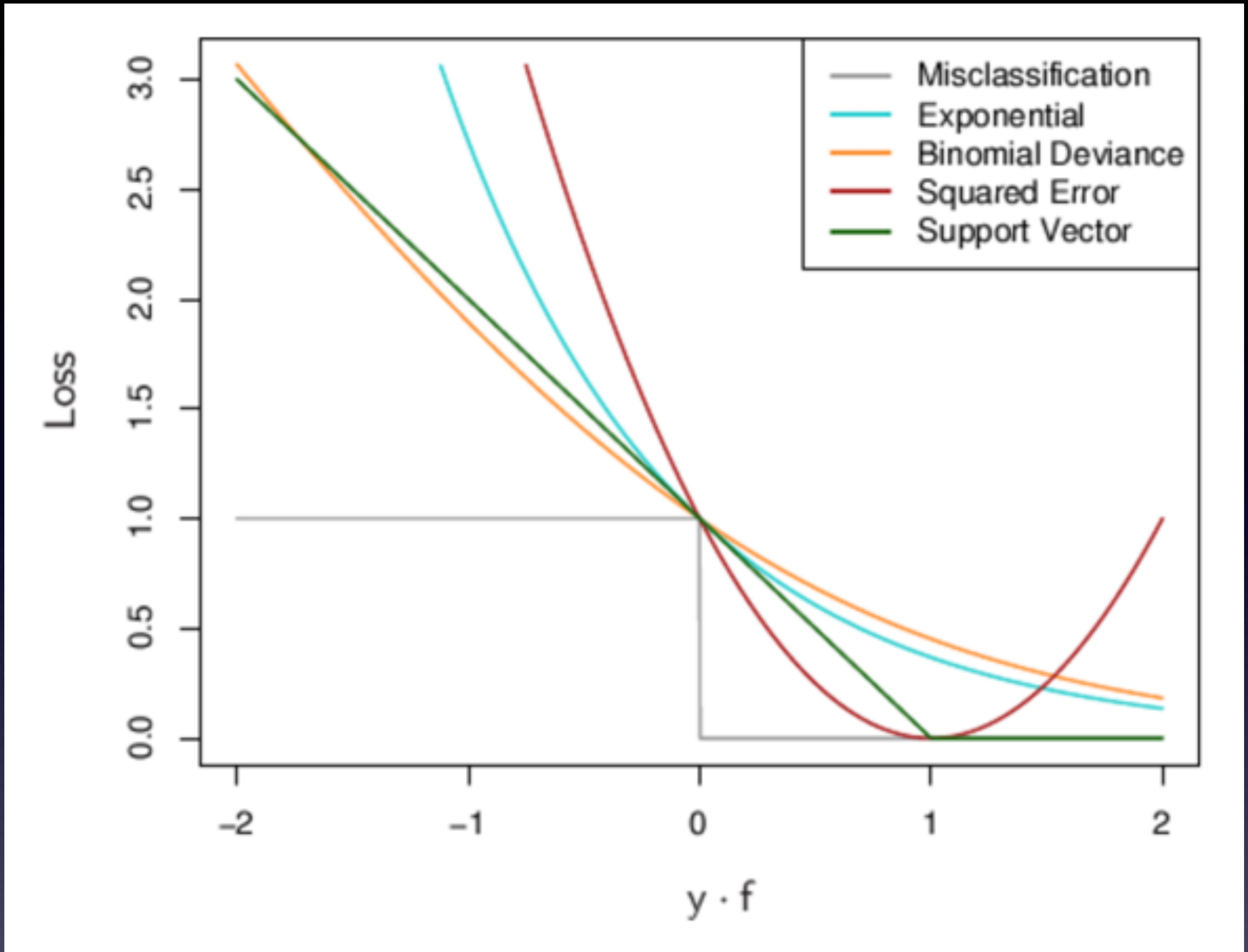


Image source: http://www.statistics4u.info/fundstat_eng/cc_data_structure.html

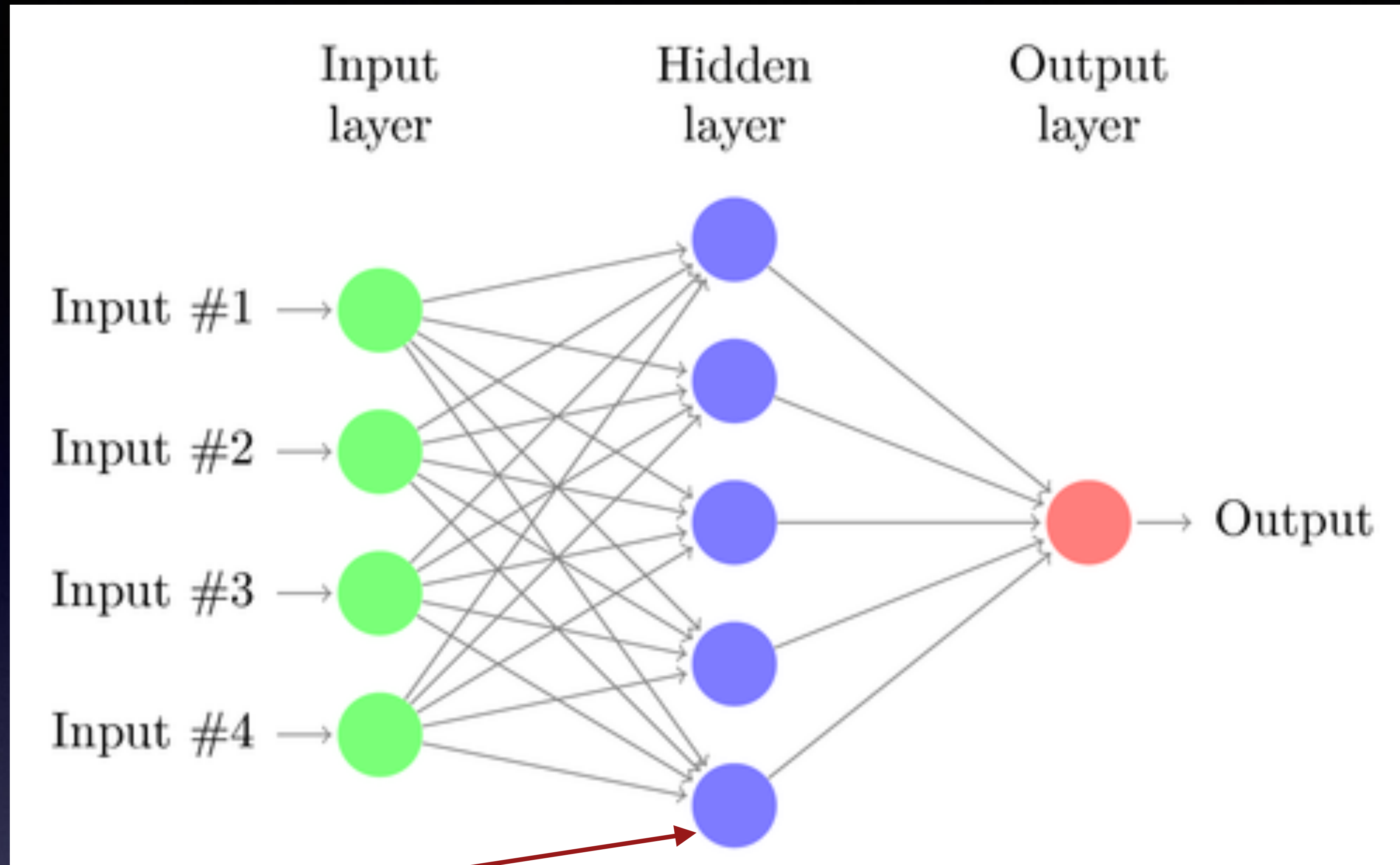
Key Ideas

- Classification
- Regression
- Features (p) and samples (n)
- Train & test data
- Loss function
- Optimization (eg. Gradient descent)



Loss functions

Image source: Elements of Statistical Learning, Hastie & Tibshirani.



$$\boldsymbol{h} = g(\boldsymbol{W}^\top \boldsymbol{x} + \boldsymbol{b})$$

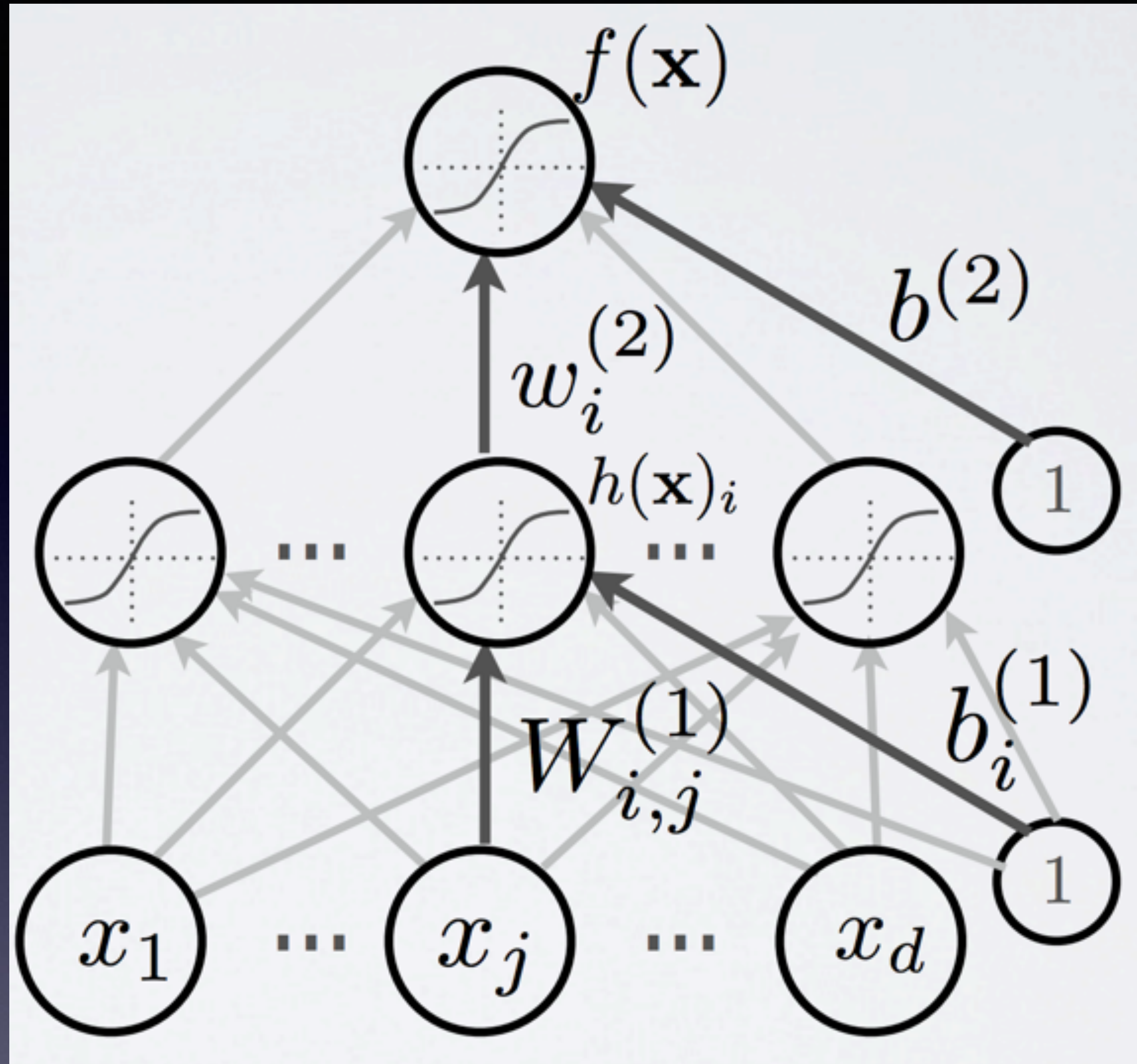
To frame as minimization, we minimize the negative log-likelihood

natural log (ln)

$$l(\mathbf{f}(\mathbf{x}), y) = -\sum_c 1_{(y=c)} \log f(\mathbf{x})_c = -\log f(\mathbf{x})_y$$

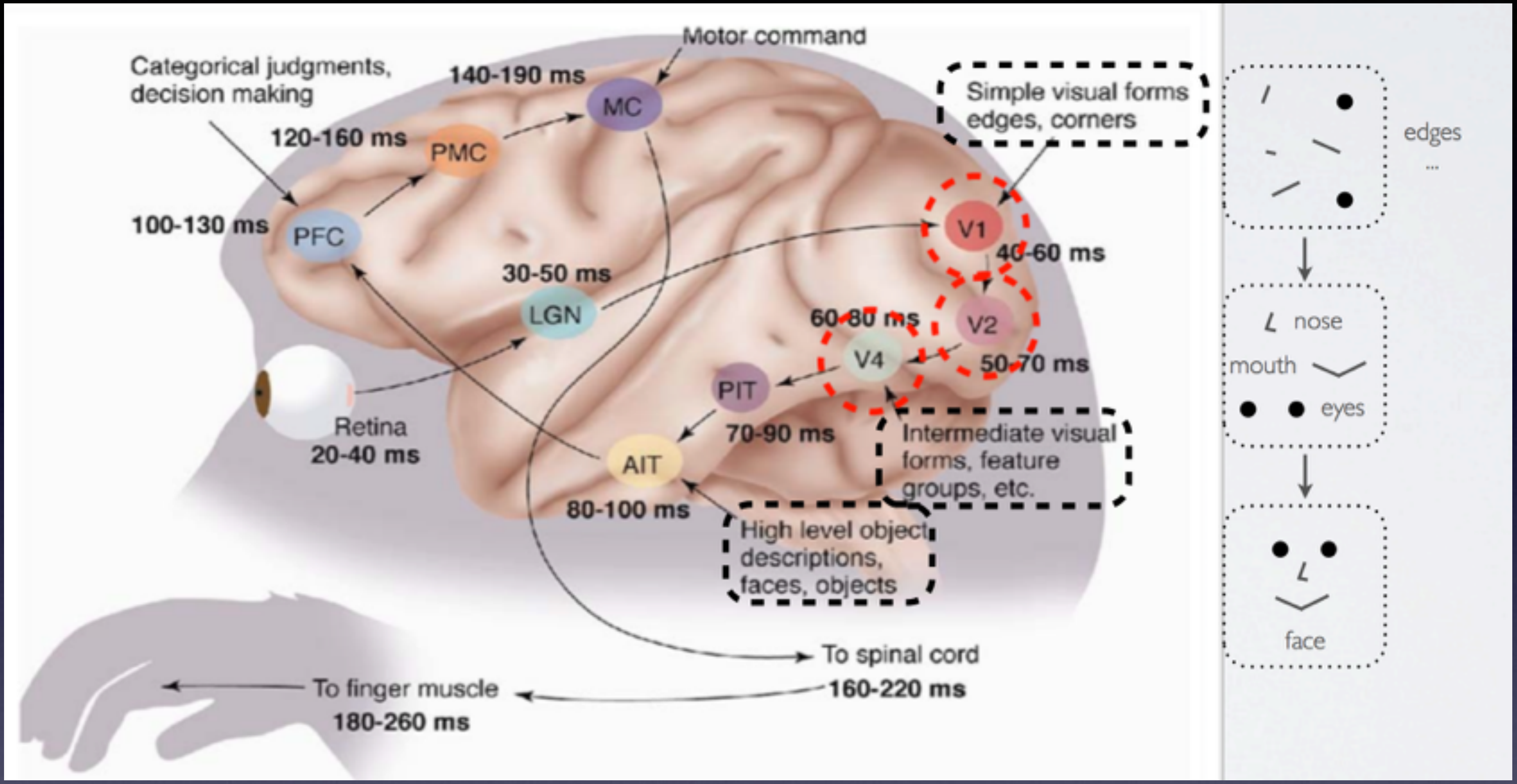
Neural Network

Image source: http://www.texample.net/tikz/examples/neural-network/, Hugo Larochelle (Neural Network slides, 2015).



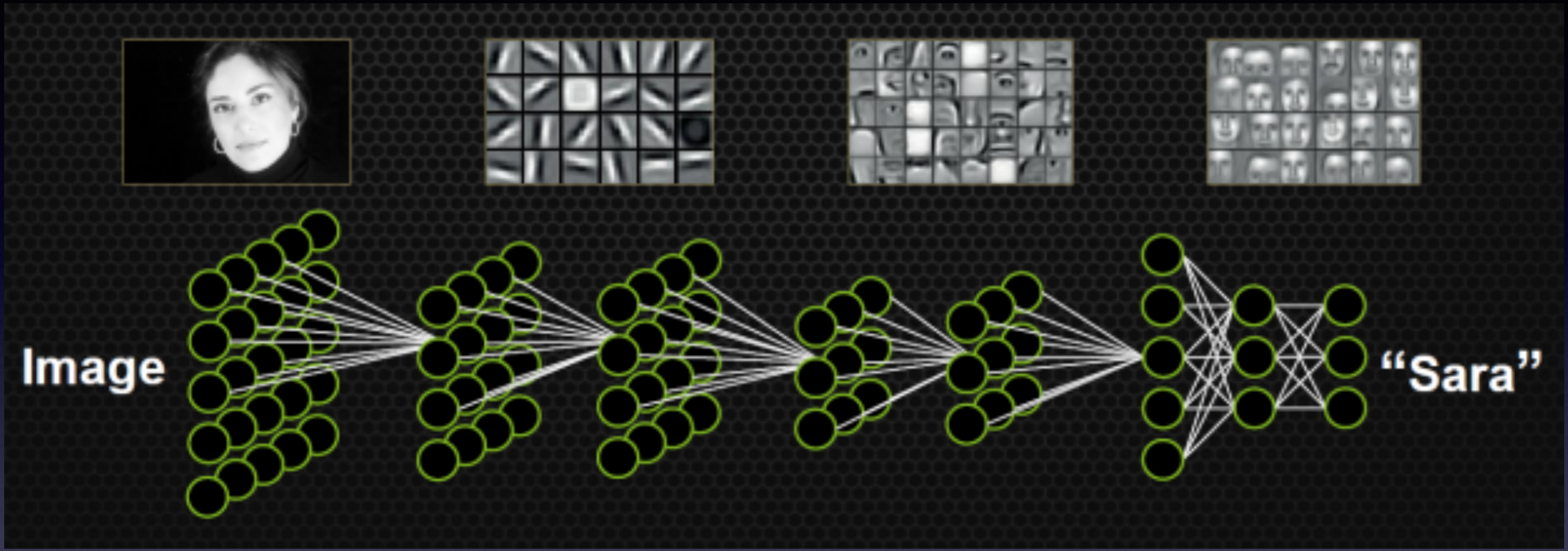
Universal approximation theorem (Hornik,1991): "a single hidden layer neural network with a linear output unit can approximate any continuous function arbitrarily well, given enough hidden units"

Image source: Hugo Larochelle, Neural Networks (2014)



Biological inspiration behind neural networks

Image source: Hugo Larochelle, Neural Networks (2014)



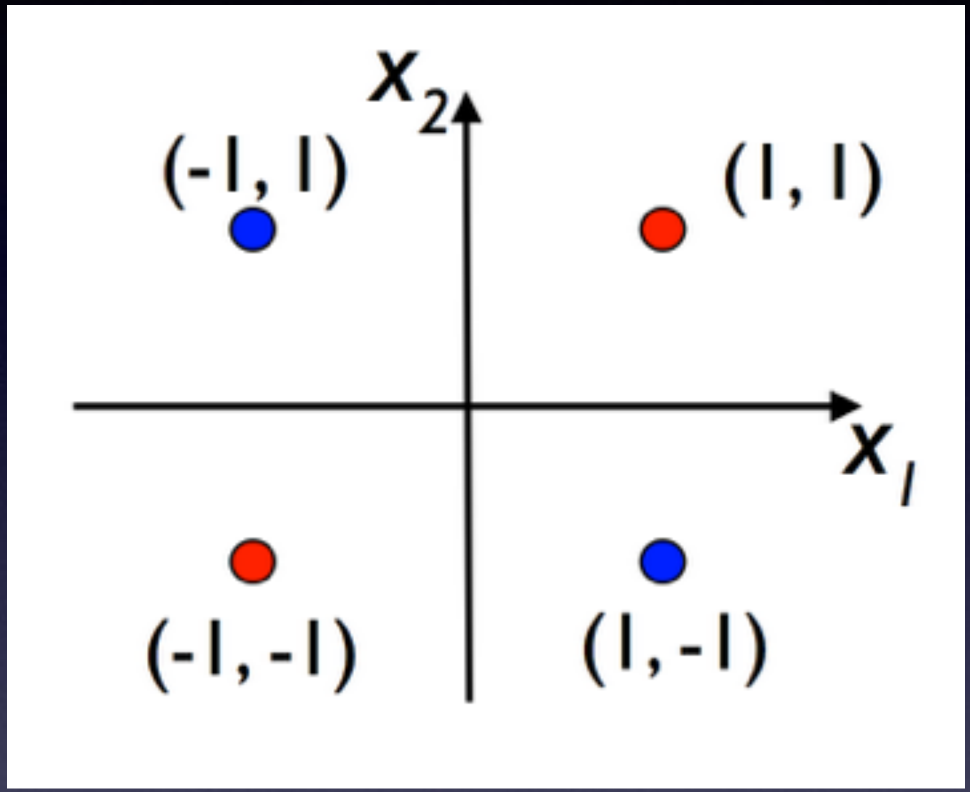
Deep Learning

Image source: https://devblogs.nvidia.com/parallelforall/accelerate-machine-learning-cudnn-deep-neural-network-library/

Learning XOR

$$f(x_1, x_2) : \{0, 1\} \rightarrow \{0, 1\}$$

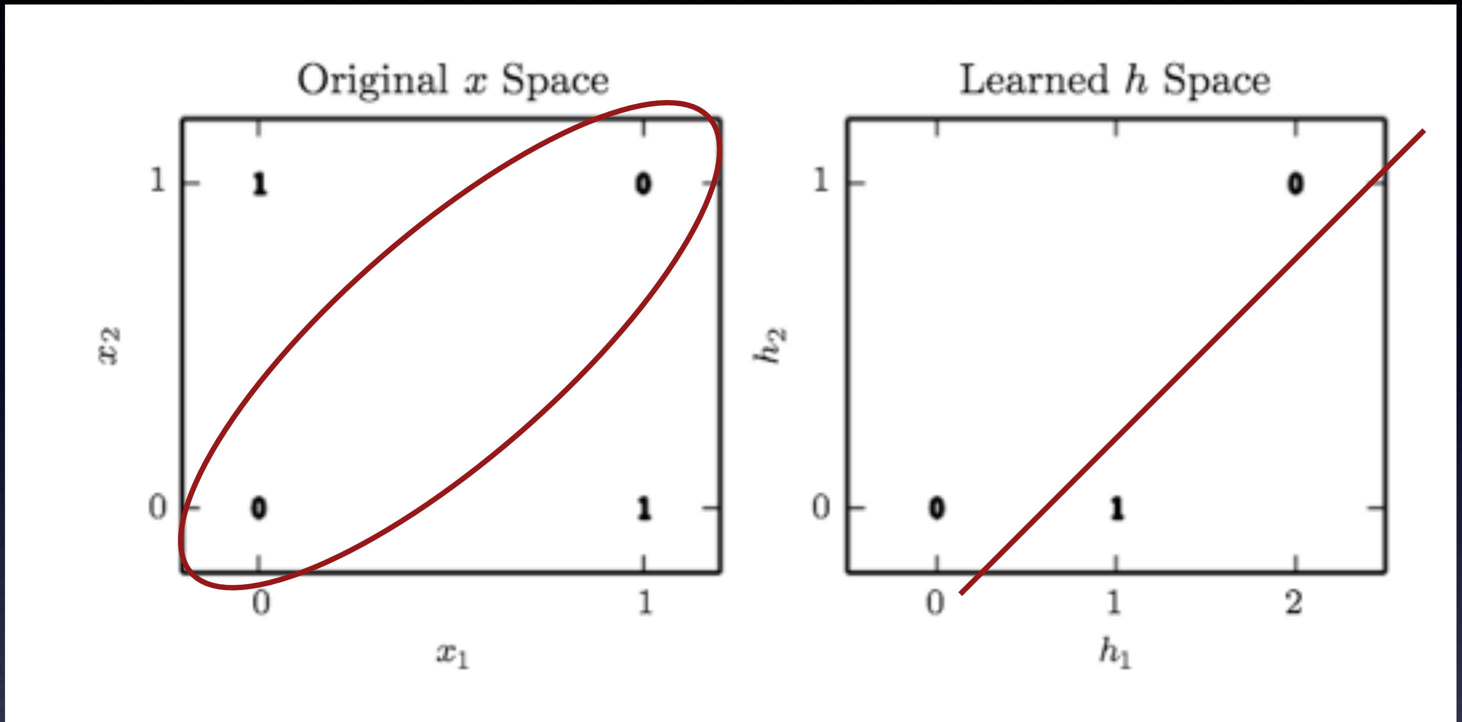
	0	**1**
0	0	1
1	1	0

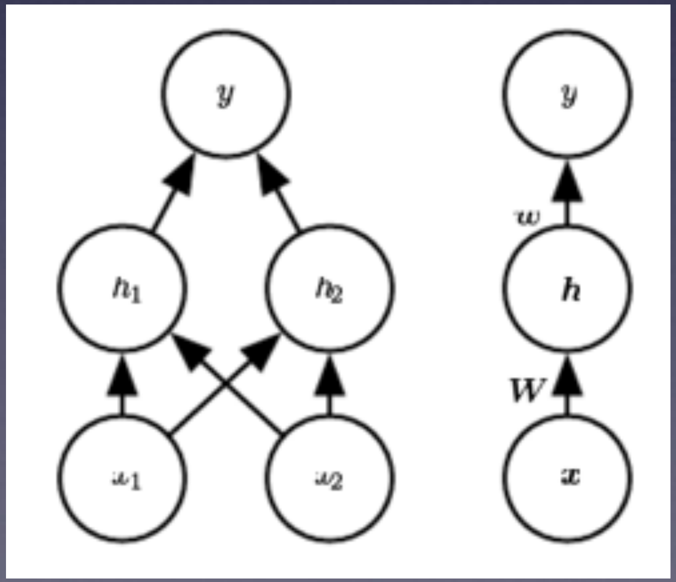


$$X = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \qquad y = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \qquad \hat{y} = ?$$

Image source: Mehryar Mohri, Foundations of Machine Learning (2015).

Learning XOR





A neural net learns using non-linearity to transform non-separable data to separable data!

Image source: http://www.deeplearningbook.org/

Gradient descent

Image source: http://eskipaper.com/amazing-mountain-valley-wallpaper.html#gal_post_26688_amazing-mountain-valley-wallpaper-1.jpg

Gradient descent for convex functions

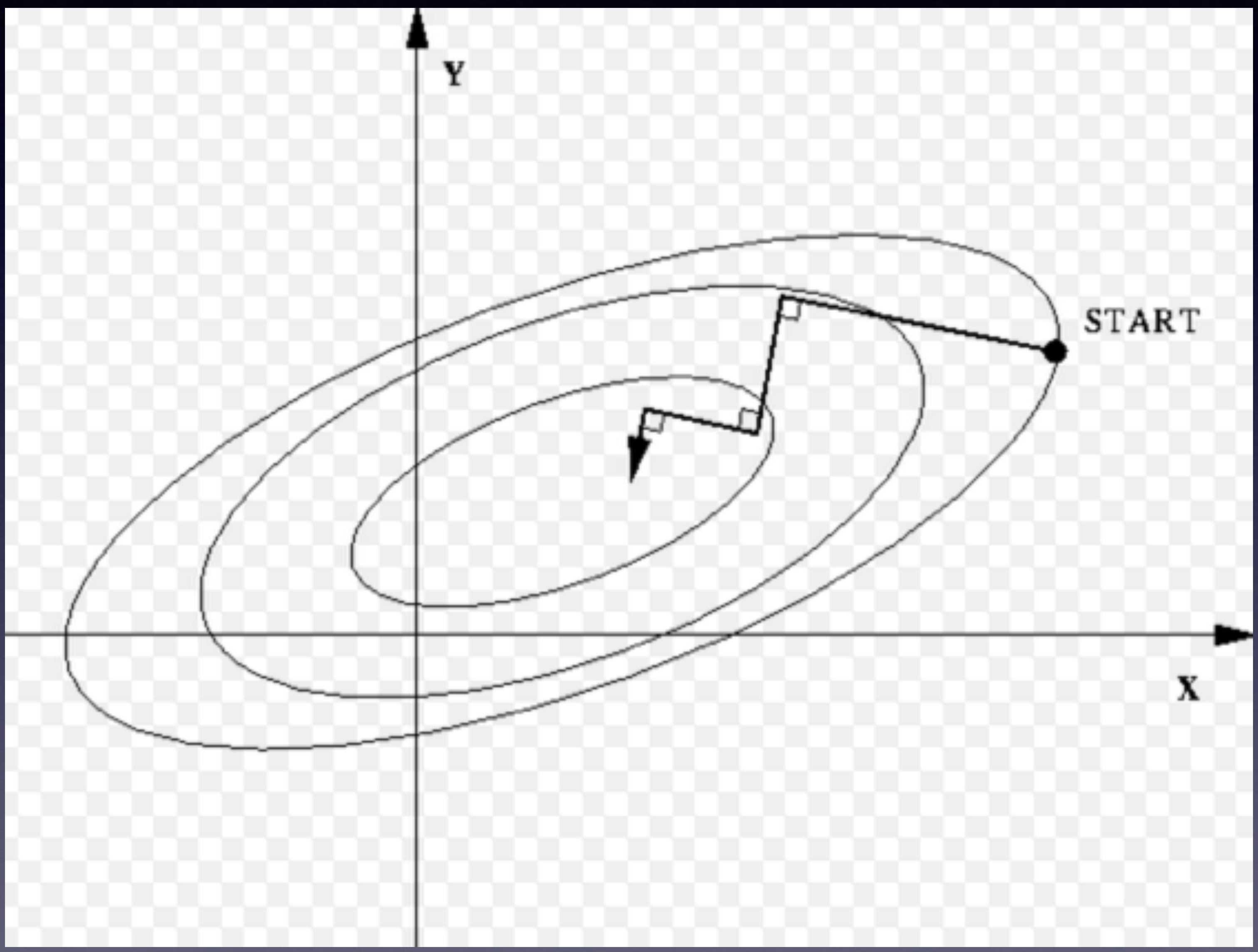
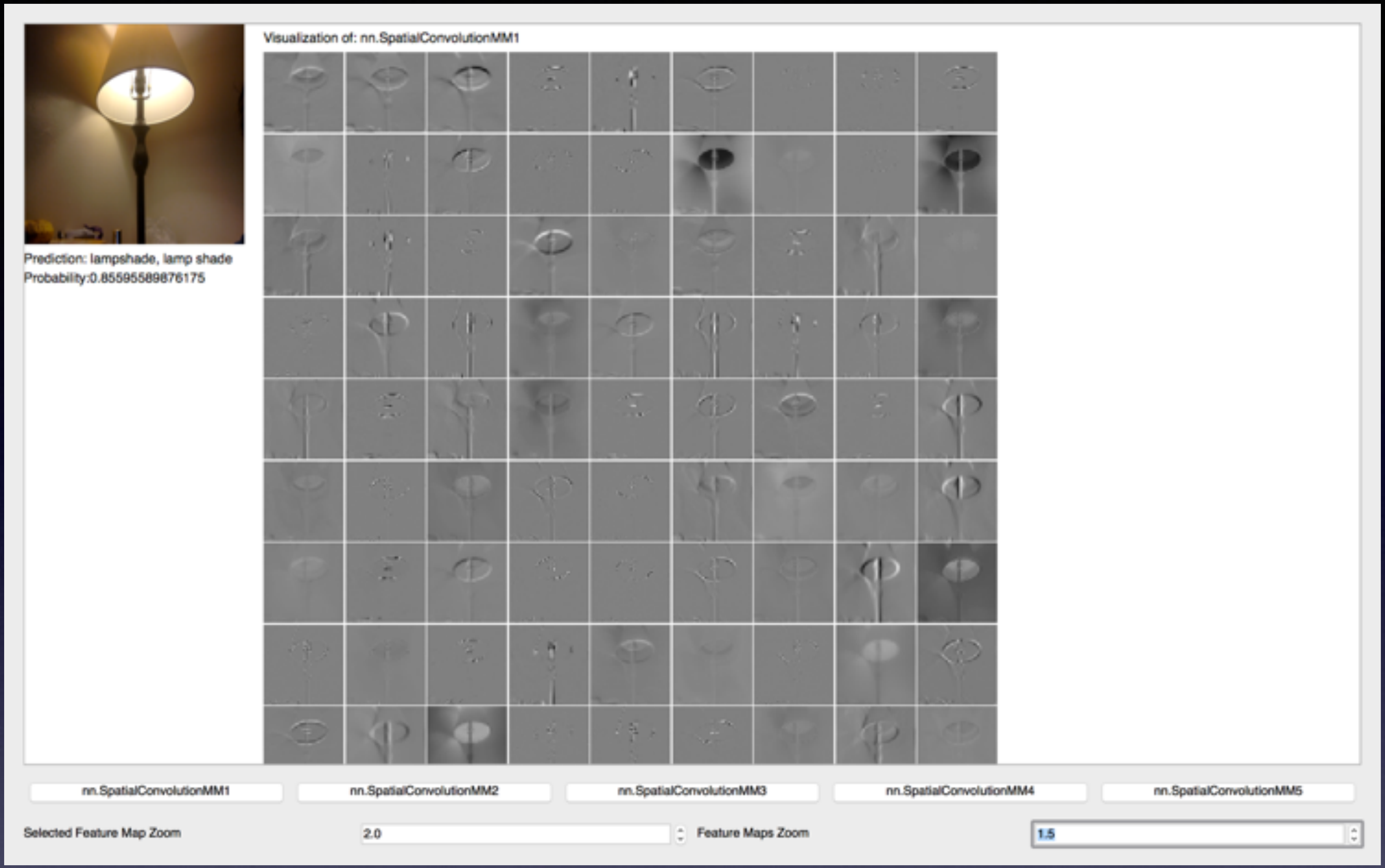


Image source: http://trond.hjorteland.com/thesis/node26.html

Stochastic gradient descent (non-convex)

Image source: https://www.reddit.com/r/EarthPorn/comments/3gspu5/a_sunset_over_the_glarus_alps_in_eastern/



ConvNet Demo

Source code: https://github.com/Aysegul/torch-visbox

Resources

1. **Machine learning**: Andrew Ng's Machine Learning class
2. **Neural networks**: Videos and slides by Hugo Larochelle
3. **Convex optimization**: MIT OCW and Boyd-Vandenberghe book
4. **Other demos**: http://cs.stanford.edu/people/karpathy/convnetjs/demo/classify2d.html, http://cs.stanford.edu/people/karpathy/convnetjs/demo/cifar10.html, http://cs.stanford.edu/people/karpathy/convnetjs/demo/mnist.html

Note: Find these slides on Github (@shivam13verma) or Twitter (@shivam13verma)