Deep Learning Theory and Applications in the Natural Sciences

Pierre Baldi

Department of Computer Science
Institute for Genomics and Bioinformatics
Center for Machine Learning and Intelligent Systems
University of California, Irvine
Scientific Discovery Drivers

1. Data (Sensors, Instruments, Data Bases, Internet, Storage…)

2. Computing (Clusters, Cloud, GPUs…)

3. Machine Learning (AI, Statistics, Data Mining, Algorithms…)

Scientific Discovery Drivers

1. **Data** (Sensors, Instruments, Data Bases, Internet, Storage…)

2. Computing (Clusters, Cloud, GPUs…)

3. Machine Learning (AI, Statistics, Data Mining, Algorithms…)

Scientific Discovery (Beginnings)
Scientific Discovery (Beginnings)
Scientific Discovery Drivers

1. Data (Sensors, Instruments, Data Bases, Internet, Storage…)
   - Exponential Growth
   - Uneveness across Fields

2. Computing (Clusters, Cloud, GPUs…)

![Image of a data center]
Scientific Discovery Drivers

1. Data (Sensors, Instruments, Data Bases, Internet, Storage…)

2. Computing (Clusters, Cloud, GPUs…)

3. Machine Learning (AI, Statistics, Data Mining, Deep Learning…)
Machine Learning: Today

Many Engineering applications (computer vision, speech recognition, NLP, robotics, e-commerce, etc.)
Simple Neuron Model

\[
\sum \begin{array}{c}
\text{Inputs} \\
\begin{array}{c}
\times_1 \\
\times_2 \\
\times_3 \\
\vdots \\
\times_n
\end{array}
\end{array} \begin{array}{c}
w_1 \\
w_2 \\
w_3 \\
w_n
\end{array} \begin{array}{c}
\text{Sum} \\
f \\
\text{Activation Function}
\end{array} \rightarrow \text{Output}
\]
Psychological Review

Theodore M. Newcomb, Editor
University of Michigan

CONTENTS

Psychological Structure and Psychological Activity .......... HELEN PEAK 325
Basic Issues in Perceptual Theory .................. W. M. O’NEIL 348
A Concept-Formation Approach to Attitude Acquisition ........ RAMON J. RHINE 362
Symptoms and Symptom Substitution ................ AUBREY J. YATES 371
Transfer of Training and Its Relation to Perceptual Learning and Recognition ........ JAMES M. VANDERPLAS 375
The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain ........ F. ROSENBLATT 386

This is the last issue of Volume 65.
Title page and index for the volume appear here.
BACK-PROPAGATION /DEEP LEARNING (~1980S)

ERROR $E(w)$

OUTPUT LAYER

$\Delta w_{ij} = \eta B_i O_j$

INPUT LAYER
Computer Vision - Image Classification

- **Imagenet**
  - Over 1 million images, 1000 classes, different sizes, avg 482x415, color
  - 16.42% Deep CNN dropout in 2012
  - 6.66% 22 layer CNN (GoogLeNet) in 2014
  - 4.9% (Google, Microsoft) super-human performance in 2015

Current error rate is below the 5% level achieved by humans.

Similar trends across engineering disciplines from speech recognition, to NLP, to games, and to robotics.

“super-human” performance in 2015
A Completely Different Style of Computation

- Storage and processing are intertwined and inseparable
- No memory addresses
- Storage is opaque: each weight contains a trace of each training example.
DL Theory

• **1-Layer Networks**
  – Perceptron theorem
  – Linear regression; Logistic regression;
  – Statistical theory and design (top layer)

• **1.5-Layer Networks**
  – Bottom layer = random  (Johnson-Lindenstrauss)
  – Bottom layer = similarities (dot products or kernels) → SVM
DL Theory

- **2-Layers Networks**
  - Universal approximation
  - Autoencoders (compressive, expansive)
  - Linear autoencoders (PCA and landscape)
  - Non-linear autoencoders (Boolean autoencoder, clustering, NP-completeness)

*FIGURE 2. The landscape of $E$.***
DL Theory

• L-Layer Networks
  – Linear
  – Boolean unrestricted
  – Local learning and its limitations (generalization of Hebb)
  – Learning Channel/Optimality of Backpropagation
  – Design (Weight sharing, Compression and Dark Knowledge, etc)
  – Dropout, Initialization, Learning rates, hyperparameter optimization

• Recurrent Networks
  – Hopfield Model and Boltzmann machines
  – Design (DAGs, LSTMs, etc)
DL Theory

• Importance of Group Theory
  – Learning permutations
  – Permutations of the units
  – Symmetries of learning rules
  – Invariant recognition
Local Deep Learning

\[ \Delta w_{ij}^{L} = F(T_i, O_i, O_{i-1}^{L}, w_{ij}^{L}) \]

\[ \Delta w_{ij}^{h} = F(I_{ij}^{h}, O_i^{h}, O_{i-1}^{h}, w_{ij}^{h}) \]

\[ \Delta w_{ij}^{1} = F(O_i, I_{ij}, w_{ij}^{1}) \]
The Learning Channel

• For optimal learning, in a physical neural system there must exist a communication channel that conveys information about the targets and the upper weights to the deep weights.
Early Beginnings

Carl Friedrich Gauss (1777 – 1855)

\[ Y = ax + b \]
\[ \Delta w_{ij}^{L} = F(T_i, O_i^L, O_{i-1}^L, w_{ij}^L) \]

\[ \Delta w_{ij}^{h} = F(I_{ij}^h, O_i^h, O_{i-1}^h, w_{ij}^h) \]

\[ \Delta w_{ij}^{1} = F(O_i, I_j, w_{ij}^1) \]
Three Fundamental Issues

• The nature of the learning channel (*Where?*).
• The semantics of the learning channel (*What?*).
• The rate of the learning channel (*How?*).
The Optimality of Backpropagation

Two Kinds of Problems and Architectures

1. Input: vectors of fixed size (e.g. computer vision). Typical architectures: feedforward neural networks.
2. Input: structured objects (e.g. graphs, molecules, sequences) of variable size. Typical architectures: recursive neural networks.

Two kinds of approaches for designing suitable networks: the inner approach and the outer approach.
The Inner and the Outer Approach for the Design of Recursive Neural Networks

1. The **Inner Approach** requires that the data and the approach be represented by a directed acyclic graph (DAG). The inner approach uses RNNs to "crawl" the edges inside the DAG.

2. The **Outer Approach** does not require a DAG. It uses RNNs in a direction that is orthogonal to (or outside) the data graph to "fold" the graph.
The Inner Approach for Sequences

DAG-HMM
The Inner Approach for Sequences

Recursive Neural Networks: DAG-RNN
The Outer Approach for Sequences

NN³

NN²

NN¹
Deep Learning in the Natural Sciences

• Physics
  - HEP: Identification of Exotic Particles (<1Å)
  - Cosmology: Identification of Quasars (10^{26}m)

• Chemistry
  - Prediction of Molecular Properties and Chemical Reactions (~1-10^2Å)

• Biology
  - Prediction of Protein Structures and Structural Features (10^2-10^4Å)

• Many more
Deep Learning in the Natural Sciences

• Physics
  - HEP: Identification of Exotic Particles (<1Å)
  - Cosmology: Identification of Quasars (10^{26}m)

• Chemistry
  - Prediction of Molecular Properties and Chemical Reactions (~1-10^{2}Å)

• Biology
  - Prediction of Protein Structures and Structural Features (10^{2}-10^{4}Å)
An important, complex, multi-faceted, somewhat ill defined problem.

Many subproblems:

- Prediction of secondary structure, relative solvent accessibility, etc.
- Prediction of tertiary structure (backbone)
- Prediction of tertiary structure (side chains).
- Prediction of quaternary structure.
Deep Learning in Biology
Deep Learning in Biology: Mining Omic Data
Progress in Accuracy (Secondary Structure)

- 60% [Chou and Fasman, 1978]
- 64% [Qian and Sejnowski, 1988]
- 74%: Rost and Sander, 1994]
- 78% [Baldi and Pollastri, 1999]
- 95% [Magnan and Baldi, 2014]
Deep Learning in Biology: Mining Omic Data

Deep Learning in Biology: Mining Omic Data

The Outer Approach for Contact Map Prediction

(In 3D, 8 hidden cubes, etc......)
2D RNNs

\[
O_{ij} = \mathcal{N}_O(I_{ij}, H_{i,j}^{NW}, H_{i,j}^{NE}, H_{i,j}^{SW}, H_{i,j}^{SE}) \\
H_{i,j}^{NE} = \mathcal{N}_E(I_{ij}, H_{i,j-1}^{NE}, H_{i,j}^{NE}) \\
H_{i,j}^{NW} = \mathcal{N}_W(I_{ij}, H_{i,j+1}^{NW}, H_{i,j-1}^{NW}) \\
H_{i,j}^{SW} = \mathcal{N}_W(I_{ij}, H_{i,j+1}^{SW}, H_{i,j-1}^{SW}) \\
H_{i,j}^{SE} = \mathcal{N}_E(I_{ij}, H_{i-1,j}^{SE}, H_{i,j+1}^{SE})
\]

The Outer Approach for Contact Map Prediction

Deep Learning


Figure 3: DEEP architecture. (a) Overview. Each **NN** represents a feed-forward neural network trainable by back-propagation. (b) For a pair of residues (i,j), the temporal input into **NN** consists of the contact probabilities produced by the network at the previous level over a neighborhood of (i,j).
Application: Breast Arterial Calcification Detection

Fig. 1 (left) A sample mammogram ROI containing BACs, (right) detection results by the deep neural network, in which detections are indicated by the green color and human marked BACs are indicated by red color.

Fig. 2. FROC curves obtained by the deep neural network.

Fig. 3. Calcium mass calculated from the human marked BACs (Reference Ca mass) and from detections (Calculated Ca mass).
Application in Drug Discovery

Healthy Vasculature

Disrupted Vasculature

ROC (area = 0.99)
Deep Learning for Drug Discovery

Healthy Vasculature

Disrupted Vasculature

ROC (area = 0.99)
Deep Learning in the Natural Sciences

- **Physics**
  - HEP: Identification of Exotic Particles (<1Å)
  - Cosmology: Identification of Quasars (10^{26}m)

- **Chemistry**
  - Prediction of Molecular Properties and Chemical Reactions (~1-10^2Å)

- **Biology**
  - Prediction of Protein Structures and Structural Features (10^2-10^4Å)
Small Molecules

**Small**
- Glycine (Gly, G) $\text{H}_2\text{N} \text{COOH}$ MW: 75.05
- Alanine (Ala, A) $\text{H}_2\text{NCH}_3\text{COOH}$ MW: 75.05
- Serine (Ser, S) $\text{H}_2\text{NCH}_2\text{COOH}$ MW: 97.09, $pK_a = 16$
- Threonine (Thr, T) $\text{H}_2\text{NCH}_2\text{CHOHCOOH}$ MW: 101.11, $pK_a = 16$
- Cysteine (Cys, C) $\text{H}_2\text{NCH}_2\text{SH}$ MW: 103.15, $pK_a = 8.35$

**Hydropathic**
- Valine (Val, V) $\text{CH}_3\text{CH}2\text{CH}_2\text{NHCOOH}$ MW: 86.14
- Leucine (Leu, L) $\text{CH}_3\text{CH}2\text{CH}2\text{CH}_2\text{NHCOOH}$ MW: 113.16
- Isoleucine (Ile, I) $\text{CH}_3\text{CH}2\text{CH}2\text{C}2\text{H}2\text{NHCOOH}$ MW: 117.16
- Methionine (Met, M) $\text{CH}_3\text{CH}2\text{CH}2\text{SCH}_3\text{NHCOOH}$ MW: 131.16
- Proline (Pro, P) $\text{CH}_2\text{CH}2\text{NHCOOH}$ MW: 97.12

**Aromatic**
- Phenylalanine (Phe, F) $\text{C}_6\text{H}_4\text{NHCOOH}$ MW: 147.18
- Tyrosine (Tyr, Y) $\text{C}_6\text{H}_4\text{OHCOOH}$ MW: 163.18
- Tryptophan (Trp, W) $\text{C}_6\text{H}_3\text{NHC}2\text{H}_4\text{COOH}$ MW: 186.21

**Acidic**
- Aspartic Acid (Asp, D) $\text{C}_2\text{H}_4\text{NO}_2$ MW: 115.09, $pK_a = 3.9$
- Glutamic Acid (Glu, E) $\text{C}_5\text{H}_9\text{NO}_4$ MW: 129.12, $pK_a = 4.67$

**Amide**
- Asparagine (Asn, N) $\text{H}_2\text{NCONH}_2$ MW: 114.11
- Glutamine (Gln, Q) $\text{H}_2\text{NCONH}_2$ MW: 126.14

**Basic**
- Histidine (His, H) $\text{H}_2\text{NCH}2\text{C}2\text{H}3\text{NH}^+$ MW: 127.14, $pK_a = 6.04$
- Lysine (Lys, K) $\text{H}_2\text{NCONH}_2$ MW: 128.17, $pK_a = 10.79$
- Arginine (Arg, R) $\text{H}_2\text{NCONH}_2$ MW: 156.19, $pK_a = 12.48$
## Chemical Space

<table>
<thead>
<tr>
<th></th>
<th>Stars</th>
<th>Small Mol.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existing</td>
<td>$10^{22}$</td>
<td>$10^7$</td>
</tr>
<tr>
<td>Virtual</td>
<td>0</td>
<td>$10^{60}$ (?)</td>
</tr>
<tr>
<td>Mode</td>
<td>Real</td>
<td>Virtual</td>
</tr>
<tr>
<td>Access</td>
<td>Difficult</td>
<td>“Easy”</td>
</tr>
</tbody>
</table>
Prediction of Molecular Properties (Physical, Chemical, Biological)

Melting Temperature? Soluble? Toxic? etc
Prediction of Molecular Properties (Physical, Chemical, Biological)

Melting Temperature? Soluble? Toxic? etc
Data Representations

NC(CO)C(=O)O

0010001001010001
Data Representations

Problem: molecular graphs are undirected

NC(CO)C(=O)O

0010001001010001
Outer Approach

Figure 1: *Left:* A visual representation of the computational graph of both standard circular fingerprints and neural graph fingerprints. First, a graph is constructed matching the topology of the molecule being fingerprinted, in which nodes represent atoms, and edges represent bonds. At each layer, information flows between neighbors in the graph. Finally, each node in the graph turns on one bit in the fixed-length fingerprint vector. *Right:* A more detailed sketch including the bond information used in each operation.

Duvenaud et al. NIPS 2015
Deep Learning in Chemistry: Predicting Chemical Reactions

RCH=CH2 + HBr → RCH(Br)–CH3

- Many important applications (synthesis, retrosynthesis, etc)
- Three different approaches:
  1. QM
  2. Rule-based systems
  3. Machine learning
Writing a System of Rules: Reaction Explorer

Table 1. SMIRKS Transformation Rules Corresponding to a Simple Alkene Hydrobromination Reaction Model

<table>
<thead>
<tr>
<th>SMIRKS</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>[C:1]=[C:2].[H:3][Cl,Br,I,S(OS=O):4]</td>
<td>alkene, protic acid addition</td>
</tr>
<tr>
<td>[H:3][C:1][C+=2].[::4]</td>
<td></td>
</tr>
<tr>
<td>[C:+1].[::2]⇒ [C+0:1][+0:2]</td>
<td>carbocation, anion addition</td>
</tr>
</tbody>
</table>

- ReactionExplorer System has about 1800 rules
- Covers undergraduate organic chemistry curriculum
- Interactive educational system
- Licensed by Wiley and distributed world-wide

Problems

- Very tedious
- Non-scalable
- Undergraduate chemistry

Table 3. Example of 10 Prioritized Transformation Rules, Relating to Alkene Hydrobromination Reactions, out of the 92 Rules Used in the Complete Robust HBr Reagent Model

<table>
<thead>
<tr>
<th>SMIRKS</th>
<th>description</th>
<th>electron flow</th>
<th>priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>[H:10][C(1:1)[C+1][C+0:2][H:10]]</td>
<td>carbocation, hydride shift from tertiary</td>
<td>1,10=10.2</td>
<td>10</td>
</tr>
<tr>
<td>[C:10][C(1:1)[C+1][C+0:2][C:10]]</td>
<td>carbocation, methyl shift from quaternary</td>
<td>1,10=10.2</td>
<td>9</td>
</tr>
<tr>
<td>[H:10][C(1:2)[C+1][C+0:2][H:10]]</td>
<td>carbocation, hydride shift from secondary</td>
<td>1,10=10.2</td>
<td>8</td>
</tr>
<tr>
<td>[C+1][C+1][C+0:2]</td>
<td>carbocation, anion addition</td>
<td>2=1</td>
<td>7</td>
</tr>
<tr>
<td>[C:1]=[C:1][C+0:2][H:3][Cl,Br,1,$(OS=O):4]$</td>
<td>alkene, protic acid addition, alkoxy</td>
<td>2,1=1,3:3:4=4</td>
<td>6</td>
</tr>
<tr>
<td>[H:3][C+1][C+2]</td>
<td>alkene, protic acid addition, benzyl</td>
<td>2,1=1,3:3:4=4</td>
<td>5</td>
</tr>
<tr>
<td>[H:3][C+1][C+2]</td>
<td>alkene, protic acid addition, allyl</td>
<td>2,1=1,3:3:4=4</td>
<td>4</td>
</tr>
</tbody>
</table>

* Included for each transformation rule is not only the SMIRKS pattern and description but also a relative priority rank to indicate the order in which the rules should be attempted. The existence of several variants for similar rules and the customized priority ordering enables robust reaction predictions that address the issues noted in Figure 7. An electron flow specification accompanies each rule to support curved arrow mechanism diagrams.
Deep Learning Chemical Reactions

RCH=CH₂ + HBr → RCH(Br)–CH₃
Siamese Architecture

source                  sink           source                 sink
Deep Learning Chemical Reactions

Figure 2. Overall reaction prediction framework: (a) A user inputs the reactants and conditions. (b) We identify potential electron donors and acceptors using coarse approximations of electron-filled and unfilled MOs. (c) Highly sensitive reactive site classifiers are trained and used to filter out the vast majority of unreactive sites, pruning the space of potential reactions. (d) Reactions are enumerated by pairing filled and unfilled MOs. (e) A ranking model is trained and used to order the reactions, where the best ranking one or few represent the major products. The top-ranked product can be recursively chained to a new instance of the framework for multistep reaction prediction.


Deep Learning in Physics

Problems at all scales from the subatomic (\(<10^{-10}\)) to the cosmological (\(10^{27}\)). Some problems (e.g. dark matter) can be approached from both scales.
Deep Learning in HEP
Standard Model Interactions
(Forces Mediated by Gauge Bosons)

- Z
  - X
  - X
  - X is any fermion in the Standard Model.

- γ
  - X
  - X
  - X is electrically charged.

- g
  - X
  - X
  - X is any quark.

- W
  - D
  - U
  - U is a up-type quark; D is a down-type quark.

- W
  - ν
  - L
  - L is a lepton and ν is the corresponding neutrino.

- W⁺
  - W⁻
  - X
  - X is a photon or Z-boson.

- W⁺
  - W⁻
  - Y
  - X and Y are any two electroweak bosons such that charge is conserved.
Deep Learning in HEP

- Higgs Boson Detection (Nature Communications, 2014)
- Supersymmetry (Nature Communications, 2014)
- Dark Knowledge (JMLR C&P 2015)
- Jet substructure, jet tagging, jet flavor, parameterized systems (Physical Review D, 2016)

- Common Features and Results:
  - dozens of features: raw + human-derived
  - millions of examples
  - classification problems
  - deep learning outperforms current methods, with or without human-derived features
Higgs Boson Detection

Simulation tools:
- MadGraph (collisions)
- PYTHIA (showering and hadronization)
- DELPHES (detector response)

11 M examples
Higgs Boson Detection

Supervised learning problem:

- Two classes
- 11 million training examples (roughly balanced)
- 28 features
  - 21 low-level features (momenta of particles)
  - 7 high-level features derived by physicists

Data available at archive.ics.uci.edu/ml/datasets/HIGGS

ML in Physics web portal: http://mlphysics.ics.uci.edu/
Higgs Boson Detection

Tuning deep neural network architectures.

<table>
<thead>
<tr>
<th>Hyper parameters</th>
<th>Choices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td>2,3,4,5,6 layers</td>
</tr>
<tr>
<td>Hidden units per layer</td>
<td>100,200,300,500</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.01, 0.05</td>
</tr>
<tr>
<td>Weight decay</td>
<td>0, 0.00001</td>
</tr>
<tr>
<td>Pre-training</td>
<td>none, autoencoder, multi-task autoencoder</td>
</tr>
<tr>
<td>Input features</td>
<td>low-level, high-level, complete set</td>
</tr>
</tbody>
</table>

Best:
- 5 hidden layers
- 300 neurons per layer
- Tanh hidden units, sigmoid output
- No pre-training
- Stochastic gradient descent
- Mini batches of 100
- Exponentially-decreasing learning rate
- Momentum increasing from .5 to .99 over 200 epochs
- Weight decay = 0.00001
Higgs Boson Detection

Deep network improves AUC by 8%

BDT = Boosted Decision Trees in TMVA package

Nature Communications, July 2014

<table>
<thead>
<tr>
<th>Technique</th>
<th>Low-level</th>
<th>High-level</th>
<th>Complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDT</td>
<td>0.73</td>
<td>0.78</td>
<td>0.81</td>
</tr>
<tr>
<td>NN</td>
<td>0.733 (0.007)</td>
<td>0.777 (0.001)</td>
<td>0.816 (0.004)</td>
</tr>
<tr>
<td>DN</td>
<td>0.880 (0.001)</td>
<td>0.800 (&lt; 0.001)</td>
<td>0.885 (0.002)</td>
</tr>
</tbody>
</table>
1. Train deep architecture using **binary targets** (multi-class classification).

2. For each training example retrieve the **soft targets** from the output of the trained deep architecture.

3. Use the soft targets (which contain the dark knowledge), to train a shallow architecture.
Dark Knowledge

Table 7: Performance of shallow networks trained with dark knowledge.

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Benchmark 1</th>
<th>Benchmark 2</th>
<th>Benchmark 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>0.842</td>
<td>0.8786</td>
<td>0.797</td>
</tr>
<tr>
<td>NN w/ dark knowledge</td>
<td>0.850</td>
<td>0.8788</td>
<td>0.799</td>
</tr>
<tr>
<td>DNN</td>
<td>0.885</td>
<td>0.8790</td>
<td>0.802</td>
</tr>
</tbody>
</table>

Jet Substructure and Jet Tagging

Figure 1: Typical jet images from class 1 (single jet) on the left, and class 2 (two overlapping jets) on the right, after preprocessing.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Signal efficiency</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDT on derived features</td>
<td>90.4%</td>
<td>96.4%</td>
</tr>
<tr>
<td>Shallow NN (32x32 input)</td>
<td>86.7% (0.01%)</td>
<td>95.2% (0.01%)</td>
</tr>
<tr>
<td>Compressed deep NN (32x32 input)</td>
<td>90.2%</td>
<td>96.3%</td>
</tr>
<tr>
<td>Deep NN (32x32 input)</td>
<td>92.7% (0.03%)</td>
<td>97.0% (0.01%)</td>
</tr>
<tr>
<td>Deep NN (48x48 input)</td>
<td>93.0%</td>
<td>97.1%</td>
</tr>
</tbody>
</table>

Table 3: Signal efficiency at 90% background rejection and AUC for each method. The best shallow neural network and the best deep neural network trained on 32 x 32 pixel images were trained with three different random initializations, and we report the mean and standard deviation for each metric.

Figure 3: ROC curves comparing the performance of a deep convolutional network trained on the images, a shallow network trained on the images, and boosted decision trees trained on the designed features.
Jet Substructure

• Use medium-level variables
  – Primary vertex (5 variables)
  – Primary tracks (8 variables)
  – Secondary vertices (5 variables)
  – Secondary tracks (8 variables)

• Use RNNS (LSTMs), CNNs,


• Neutrino experiments
• Antimatter experiments
Neural Networks Criticisms

- Black-magic (design and hyperparameter tuning, etc)
- Black-box
Neural Networks Criticisms

• Black-magic (design and hyperparameter tuning, etc)
• Black-box
  – Reasonable concern
Neural Networks Criticisms

- Black-magic (design and hyperparameter tuning, etc)
- Black-box
  - Reasonable and very human concern
  - Black box can be opened (can be interesting but expensive)
Neural Networks Criticisms

• Black-magic (design and hyperparameter tuning, etc)
• Black-box
  – Reasonable and very human concern
  – Black box can be opened (can be interesting but expensive)
  – A certain degree of opacity is to be expected (non-linear surfaces in high dimensional spaces, loading of training examples into the weights)
Neural Networks Criticisms

• Black-magic (design and hyperparameter tuning, etc)
• Black-box
  – Reasonable and very human concern
  – Black box can be opened (can be interesting but expensive)
  – A certain degree of opacity is to be expected (non-linear surfaces in high dimensional spaces, loading of training examples into the weights)
  – Problem of performance guarantees
Neural Networks Criticisms

• Black-magic (design and hyperparameter tuning, etc)
• Black-box
  – Reasonable and very human concern
  – Black box can be opened (can be interesting but expensive)
  – A certain degree of opacity is to be expected (non-linear surfaces in high dimensional spaces, loading of training examples into the weights)
  – Problem of performance guarantees
  – Greatest irony of all: the brain is a black box
THANK YOU

The detailed, all-sky picture of the infant universe. The image reveals 13.77 billion year old temperature fluctuations (shown as color differences) that correspond to the seeds that grew to become the galaxies. Credit: NASA / WMAP Science Team WMAP # 121238 Image Caption 9 year WMAP image of background cosmic radiation (2012)