INTRODUCTION TO DEEP LEARNING

Steve Tjoa
kiemyang@gmail.com

June 2013

Acknowledgements

- http://ufldl.stanford.edu/wiki/index.php/UFLDL_Tutorial
- http://youtu.be/AyzOUbkUf3M
- http://youtu.be/ZmNOAtZIgIk

What is Deep Learning?

- "a class of machine learning techniques, developed mainly since 2006, where many **layers** of **non-linear** information processing stages or **hierarchical architectures** are exploited."
- "recently applied to many signal processing areas such as image, video, audio, speech, and text and has produced surprisingly good results"
  - http://www.icassp2012.com/Tutorial_09.asp

- “technology companies are reporting startling gains in fields as diverse as computer vision, speech recognition and the identification of promising new molecules for designing drugs”
- “has already been put to use in services like Apple’s Siri virtual personal assistant, which is based on Nuance Communications’ speech recognition service, and in Google’s Street View, which uses machine vision to identify specific addresses”
  - http://www.nytimes.com/2012/11/24/science/scientists-see-advances-in-deep-learning-a-part-of-artificial-intelligence.html?hpw&pagewanted=all

The New York Times

Science

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

ENVIRONMENT SPACE & COSMOS

Scientists See Promise in Deep-Learning Programs

Hao Zhang/The New York Times

A voice recognition program translated a speech given by Richard F. Rashid, Microsoft's top scientist, into Mandarin Chinese.

By JOHN MARKOFF
Published: November 23, 2012

A Brief History

- 1950s: Artificial neural networks mimic the way the brain absorbs information and learns from it.
- 1960s: computer scientists: “a workable artificial intelligence system is just 10 years away!”
- 1980s: a wave of commercial start-ups collapsed, leading to what some people called the “A.I. winter.”
- 1990s: SVMs!

Reducing the Dimensionality of Data with Neural Networks

G. E. Hinton* and R. R. Salakhutdinov

Keith Penner

A student team led by the computer scientist Geoffrey E. Hinton used deep-learning technology to design software.

- **2006**: Geoffrey Hinton pioneers powerful new techniques for helping the artificial networks recognize patterns.

- 2006–present: Andrew Ng and others help popularize the method.
- 2013: Google acquires Hinton's deep learning startup.

Why Neural Networks?

- People are better than computers at recognizing patterns.
- Neurons in the perceptual system represent features of sensory input.
- The brain learns **layers** of features.

Why So Popular?

- **Scalable**. "...it scales beautifully. Basically you just need to keep making it bigger and faster, and it will get better." ~Hinton
- **Accurate**. Jeff Dean and Andrew Ng "programmed a cluster of 16,000 computers to train itself to automatically recognize images in a library of 14 million pictures of 20,000 different objects. ... the system did 70 percent better than the most advanced previous one."

- A lab at the University of Lugano "won a pattern recognition contest by outperforming both competing software systems and a human expert in identifying images in a database of German traffic signs."
- "The winning program accurately identified 99.46 percent of the images in a set of 50,000; the top score in a group of 32 human participants was 99.22 percent, and the average for the humans was 98.84 percent."

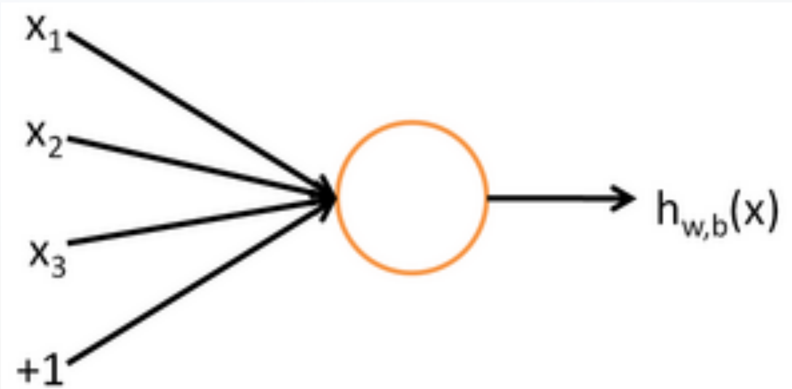
REWIRING CORTEX: FUNCTIONAL PLASTICITY OF THE AUDITORY CORTEX DURING DEVELOPMENT

Jessica R. Newton and Mriganka Sur

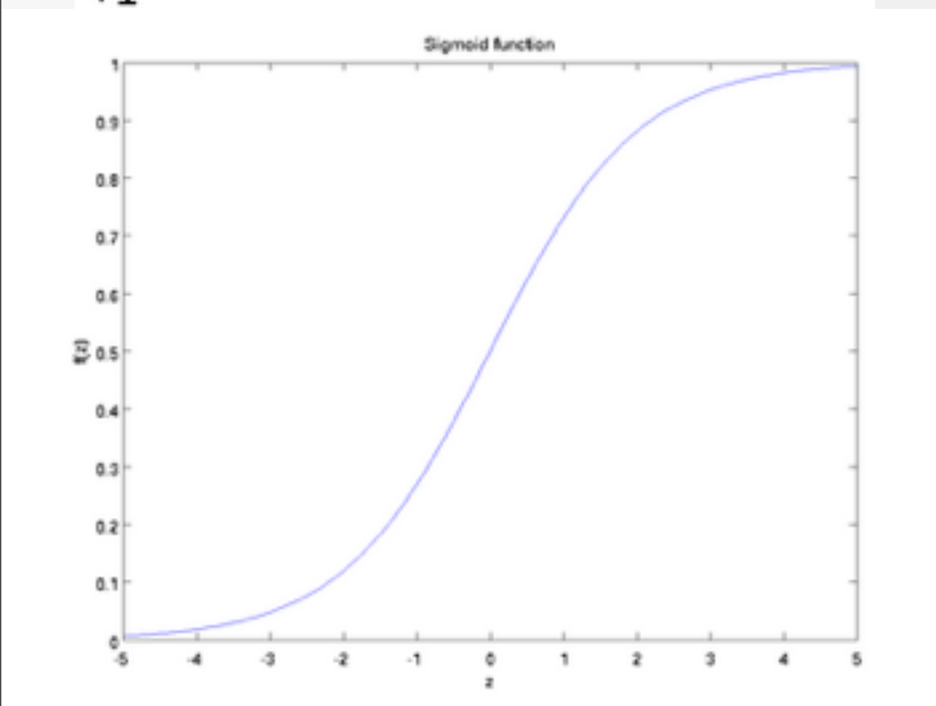
Department of Brain & Cognitive Sciences, Picower Center for Learning & Memory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA

- **Adaptive.** In general, early on, neurons are not function specific. The auditory cortex can learn to see!

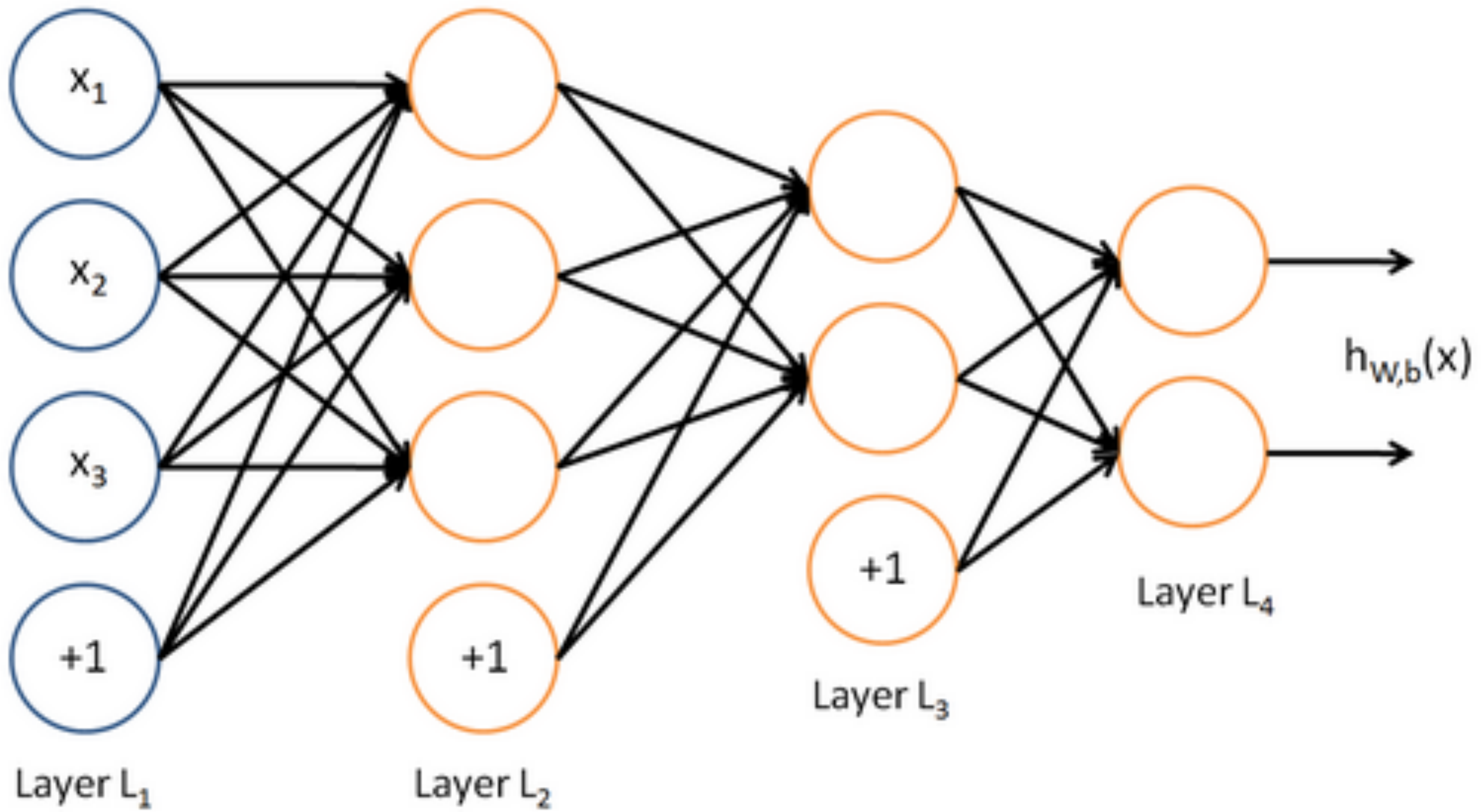
Basic Concepts



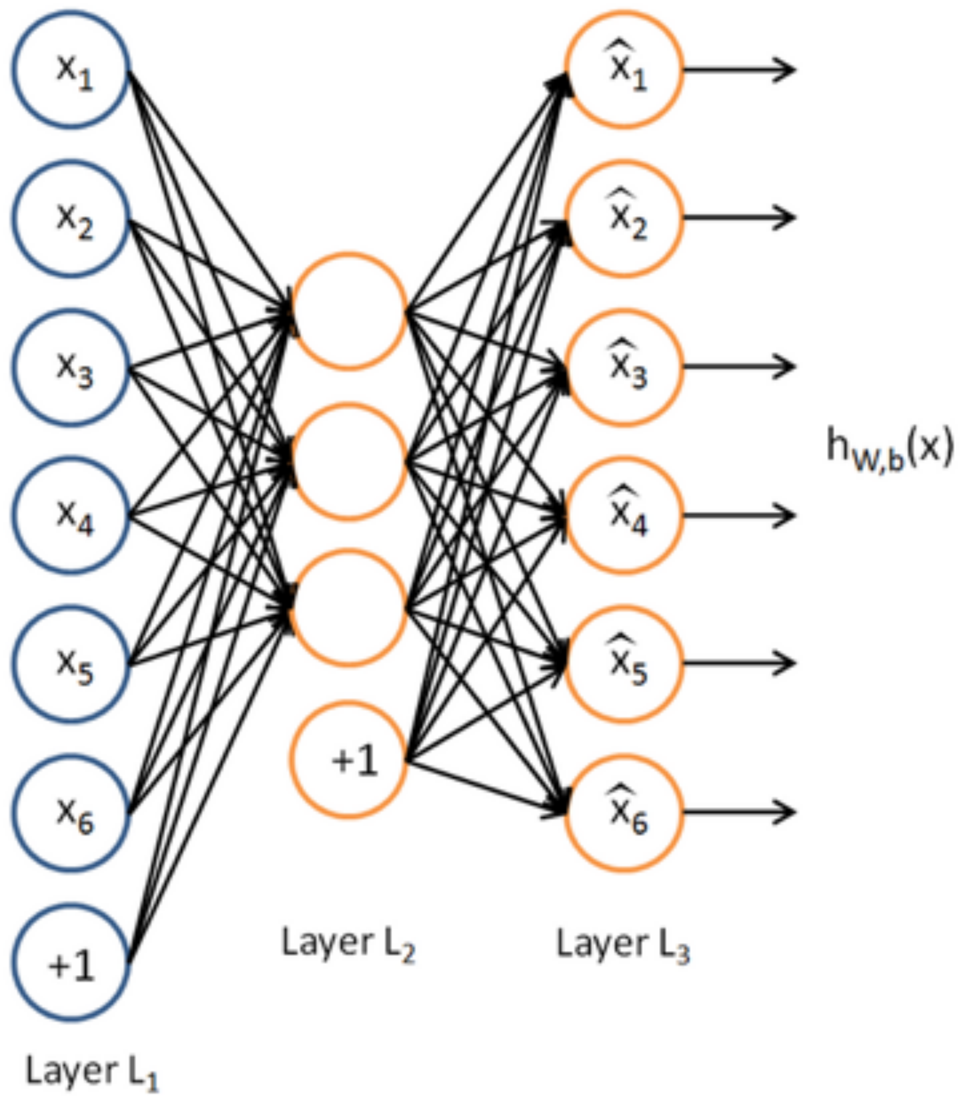
- Neuron:
  $h(x) = f(w^T x + b)$
- Parameters to train: w and b



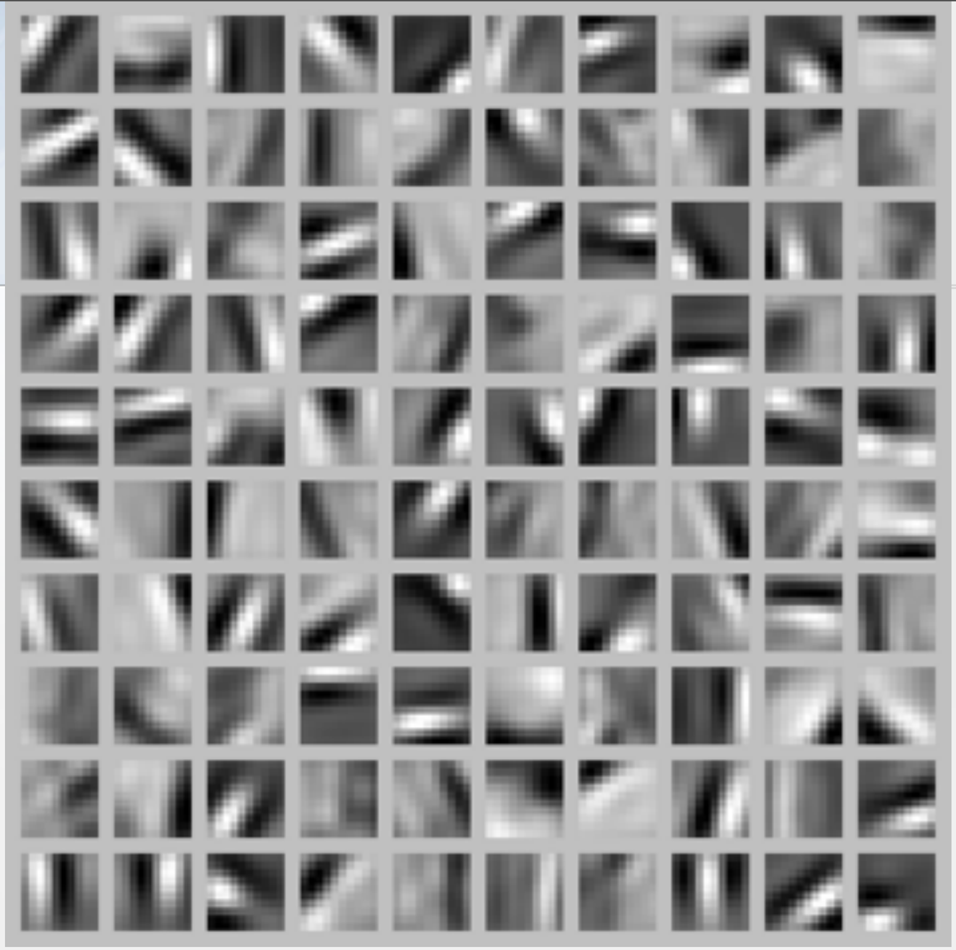
$$f(z) = \frac{1}{1 + \exp(-z)}.$$



- Stack layers of neurons.
- Problem: given input, x, and output, y, find parameters, w.
- Training algorithm: **back propagation**.



- **Autoencoder**: a special kind of NN
- input layers and output layers are equal



- Example autoencoder: 10-by-10 pixel images, and 100 hidden units

Self-Taught Learning

- Use the learned activations as features.
- http://ufldl.stanford.edu/wiki/index.php/Self-Taught_Learning

Deep Networks

- Many layers can model more complex features than few layers.
- **Difficulty**: training!
- Solution: greedy layer–wise training.
- Restricted Boltzmann Machine (RBM)
- Contrastive Divergence (CD)

Building High-level Features Using Large Scale Unsupervised Learning

Quoc V. Le QUOCLE@CS.STANFORD.EDU
Marc'Aurelio Ranzato RANZATO@GOOGLE.COM
Rajat Monga RAJATMONGA@GOOGLE.COM
Matthieu Devin MDEVIN@GOOGLE.COM
Kai Chen KAICHEN@GOOGLE.COM
Greg S. Corrado GCORRADO@GOOGLE.COM
Jeff Dean JEFF@GOOGLE.COM
Andrew Y. Ng ANG@CS.STANFORD.EDU

- ICML 2012
- Traditional ML model: feature extraction, then (supervised) machine learning.
- Instead: **learn** good features, then **cluster** them.

Deep learning with COTS HPC systems

Adam Coates ACOATES@CS.STANFORD.EDU
Brody Huval BRODYH@STANFORD.EDU
Tao Wang TWANGCAT@STANFORD.EDU
David J. Wu DWU4@CS.STANFORD.EDU
Andrew Y. Ng ANG@CS.STANFORD.EDU
Stanford University Computer Science Dept., 353 Serra Mall, Stanford, CA 94305 USA

Bryan Catanzaro BCATANZARO@NVIDIA.COM
NVIDIA Corporation, 2701 San Tomas Expressway, Santa Clara, CA 95050

- ICML 2013
- Training a huge system is overwhelming!
- Proposes a deep belief network built with a GPU cluster and commodity hardware.

Unsupervised feature learning for audio classification using convolutional deep belief networks

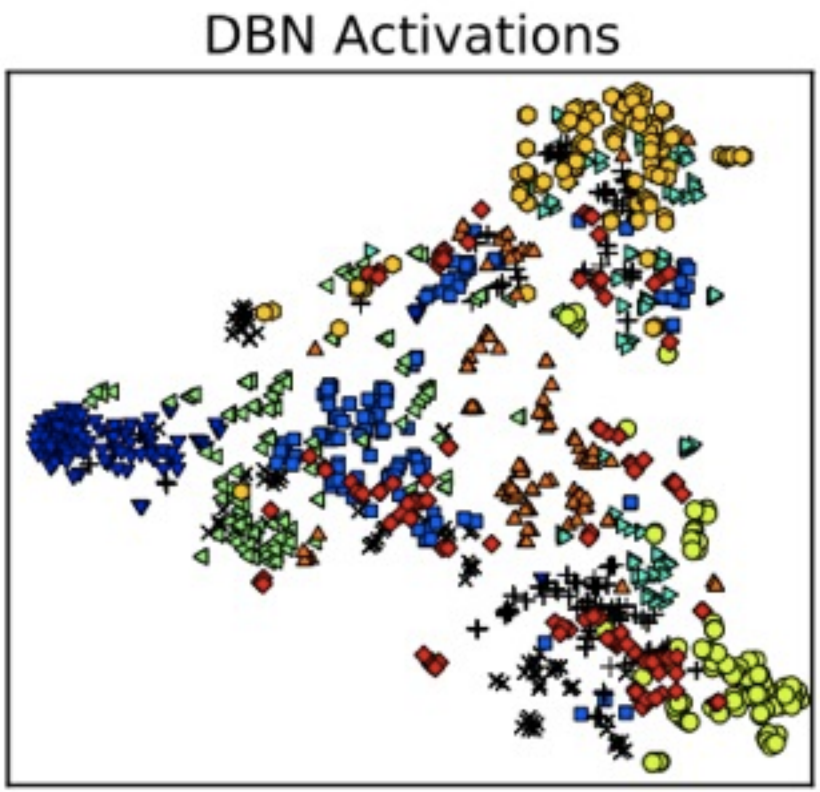
Honglak Lee **Yan Largman** **Peter Pham** **Andrew Y. Ng**
Computer Science Department
Stanford University
Stanford, CA 94305

- NIPS 2009
- For speech: speaker recognition, gender recognition, phoneme recognition
- For music: genre recognition, artist recognition
  - Just give it the spectrogram!

11th International Society for Music Information Retrieval Conference (ISMIR 2010)

LEARNING FEATURES FROM MUSIC AUDIO WITH DEEP BELIEF NETWORKS

Philippe Hamel and Douglas Eck
DIRO, Université de Montréal
CIRMMT
{hamelphi,eckdoug}@iro.umontreal.ca



- SVM with RBF upon the output activations
  - outperforms MFCCs
- genre recognition, autotagging
- “there are many hyper-parameters to optimize”

A CLASSIFICATION-BASED POLYPHONIC PIANO TRANSCRIPTION APPROACH USING LEARNED FEATURE REPRESENTATIONS

Juhan Nam
Stanford University
juhan@ccrma.stanford.edu

Jiquan Ngiam
Stanford University
jngiam@cs.stanford.edu

Honglak Lee
Univ. of Michigan, Ann Arbor
honglak@eecs.umich.edu

Malcolm Slaney
Yahoo! Research
malcolm@ieee.org

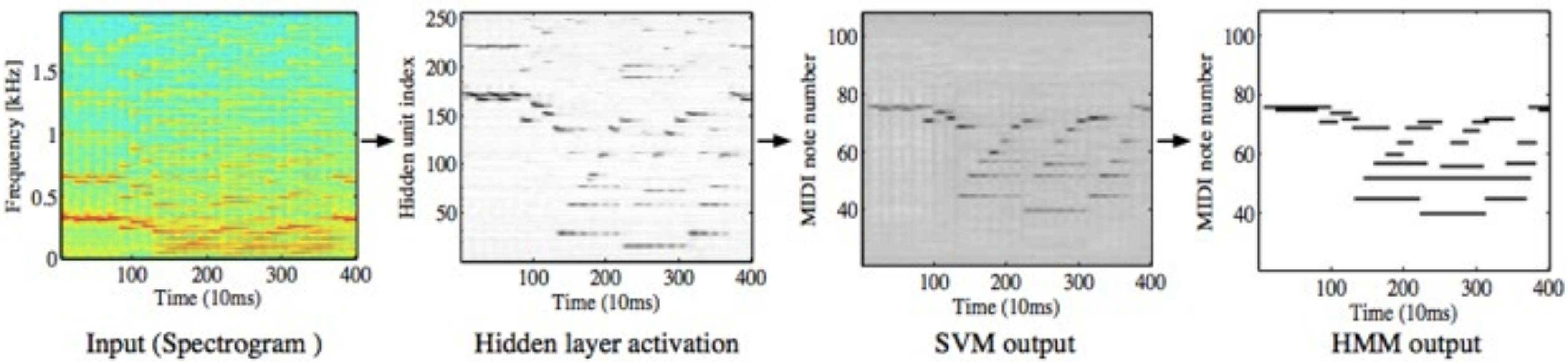


Figure 3: Signal transformation through the DBNs and classification stages

- ISMIR 2011

12th International Society for Music Information Retrieval Conference (ISMIR 2011)

AUDIO-BASED MUSIC CLASSIFICATION WITH A PRETRAINED CONVOLUTIONAL NETWORK

Sander Dieleman, Philémon Brakel and Benjamin Schrauwen
Electronics and Information Systems department, Ghent University
{sander.dieleman, philemon.brakel, bschrauw}@elis.ugent.be

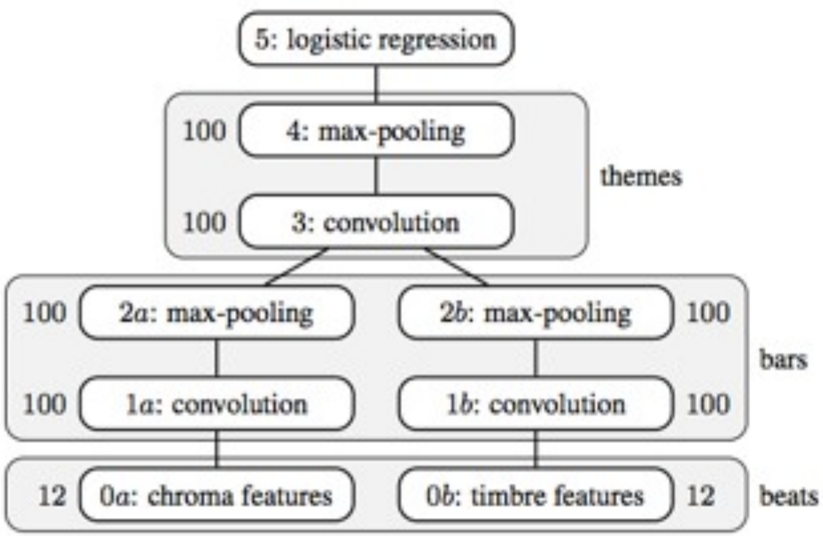


Figure 3. The network layout. The number of dimensions or feature maps for each layer is indicated on the side. The layers have also been grouped according to the timescale on which they operate.

- artist recognition, genre recognition, key detection
- on the Million Song Dataset

13th International Society for Music Information Retrieval Conference (ISMIR 2012)

ANALYZING DRUM PATTERNS USING CONDITIONAL DEEP BELIEF NETWORKS

Eric Battenberg
University of California, Berkeley
Dept. of Electrical Engineering and Computer Sciences
ericb@eecs.berkeley.edu

David Wessel
University of California, Berkeley
Center for New Music and Audio Technologies
wessel@cnmat.berkeley.edu

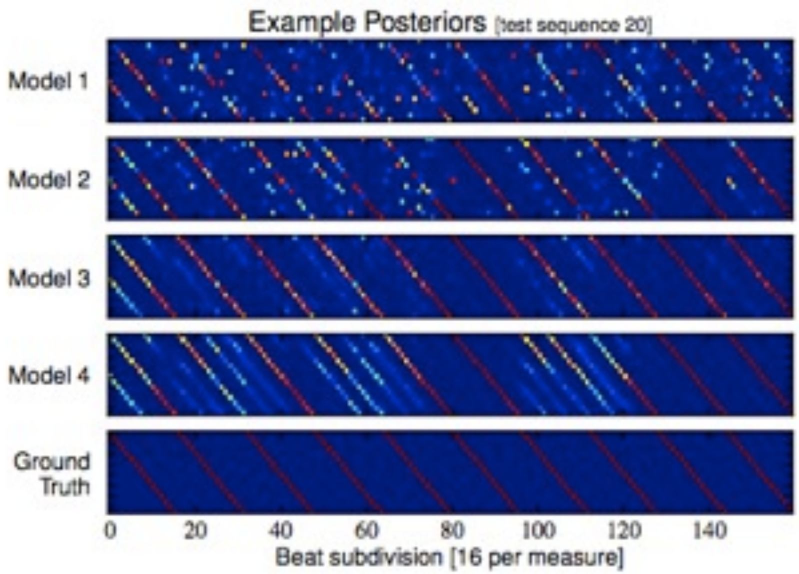


Figure 6. Example posteriors subdivision probabilities from the four models and the ground truth labels. The columns in each matrix show the posterior probability of each label for a particular beat subdivision.

- Goal: "identifying the alignment of beats within a measure"
- Features: drum onset patterns (bounded linear units)