Deep Learning and Speech Processing
An Introduction

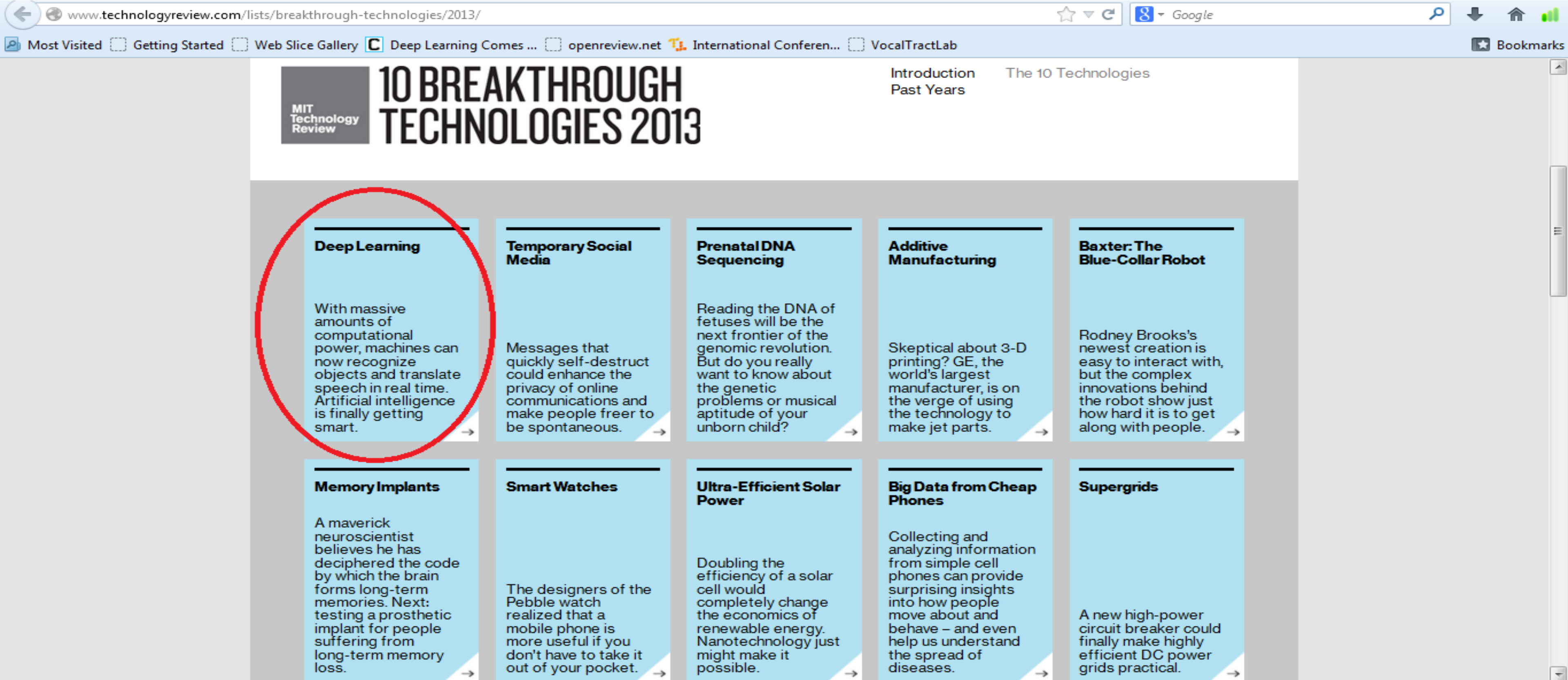
By Hazrat Ali
(Pakistan)

Outline

- Motivation
- Strength of Deep Learning
- Deep Architecture
- Deep Belief Network
- Our Approach
- Example
- Resources

Motivation

- Deep Learning is on the top of MIT Technology Review Breakthrough Technologies of 2013
- Ref http://www.technologyreview.com/lists/breakthrough-technologies/2013/



10 BREAKTHROUGH TECHNOLOGIES 2013

Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.

Temporary Social Media

Messages that quickly self-destruct could enhance the privacy of online communications and make people freer to be spontaneous.

Prenatal DNA Sequencing

Reading the DNA of fetuses will be the next frontier of the genomic revolution. But do you really want to know about the genetic problems or musical aptitude of your unborn child?

Additive Manufacturing

Skeptical about 3-D printing? GE, the world's largest manufacturer, is on the verge of using the technology to make jet parts.

Baxter: The Blue-Collar Robot

Rodney Brooks's newest creation is easy to interact with, but the complex innovations behind the robot show just how hard it is to get along with people.

Memory Implants

A maverick neuroscientist believes he has deciphered the code by which the brain forms long-term memories. Next: testing a prosthetic implant for people suffering from long-term memory loss.

Smart Watches

The designers of the Pebble watch realized that a mobile phone is more useful if you don't have to take it out of your pocket.

Ultra-Efficient Solar Power

Doubling the efficiency of a solar cell would completely change the economics of renewable energy. Nanotechnology just might make it possible.

Big Data from Cheap Phones

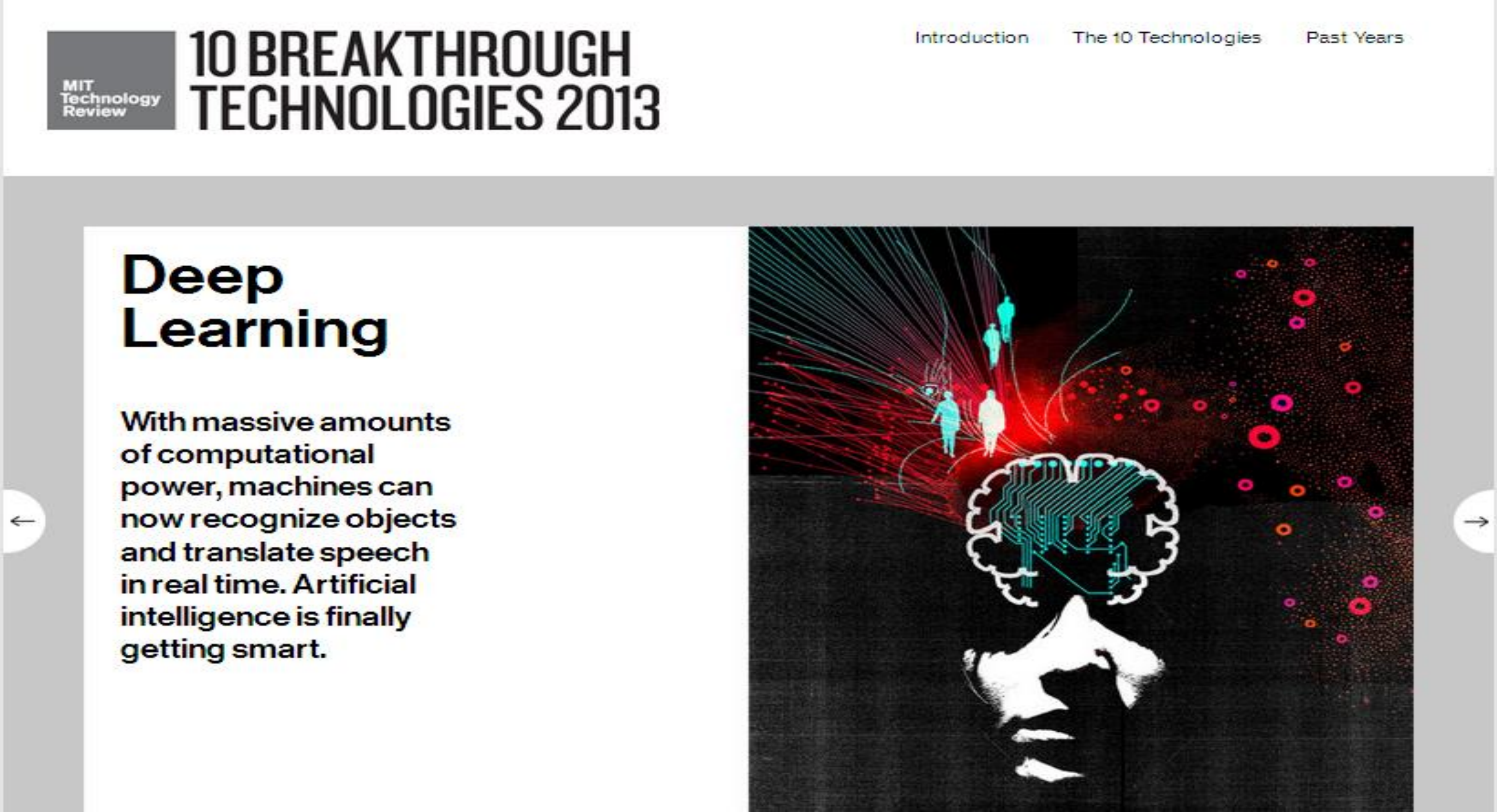
Collecting and analyzing information from simple cell phones can provide surprising insights into how people move about and behave – and even help us understand the spread of diseases.

Supergrids

A new high-power circuit breaker could finally make highly efficient DC power grids practical.

Motivation

- Artificial Intelligence getting smarter
- Ref: http://www.technologyreview.com/featuredstory/513696/deep-learning/



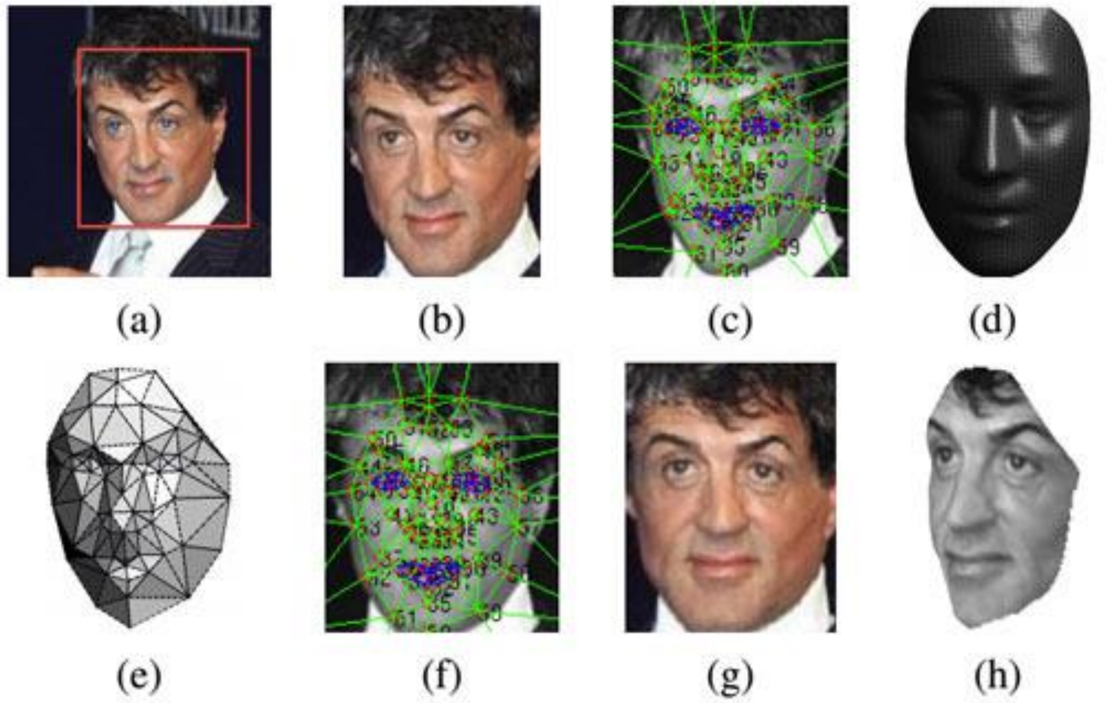
10 BREAKTHROUGH TECHNOLOGIES 2013

Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.

Motivation

- **DeepFace** Project by Facebook
  - Closely matches human performance for face recognition
  - http://www.technologyreview.com/news/525586/facebook-creates-software-that-matches-faces-almost-as-well-as-you-do/



Motivation

- **Google buys DeepMind**
  - DeepMind is a UK based Artificial Intelligence startup
  - Less than a dozen engineers
    - Why did Google pay such a huge amount for a small company?

400 Million Pounds

≈

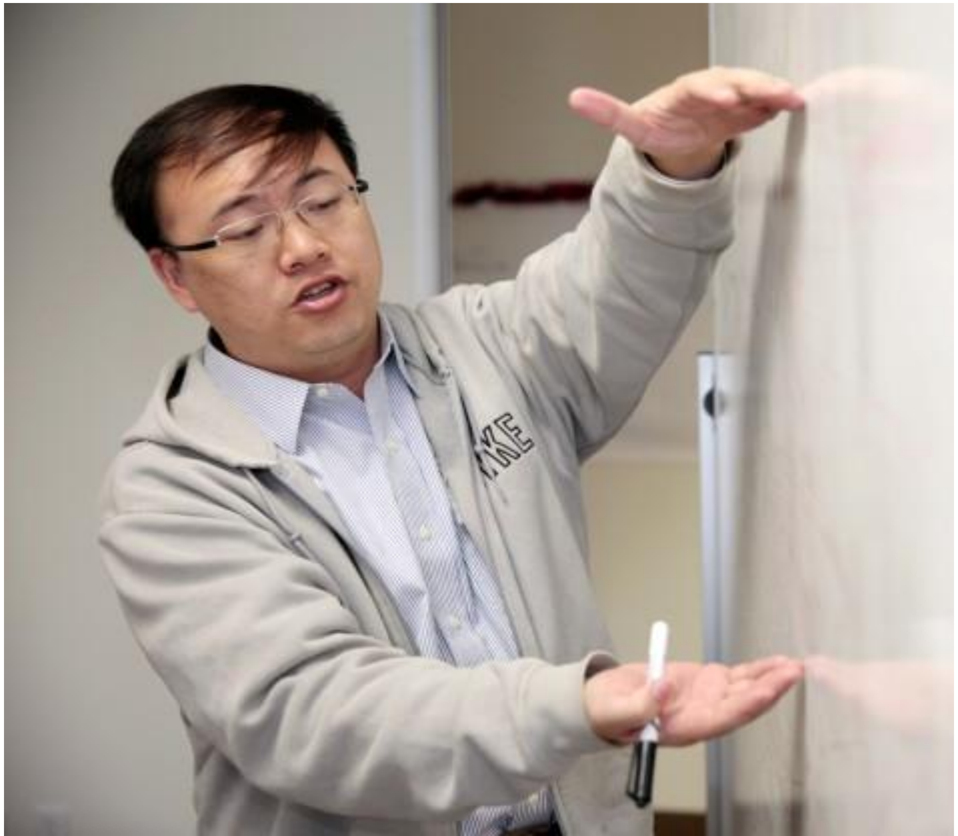
4,000,000,000 ￥

- http://www.theguardian.com/technology/2014/jan/27/google-acquires-uk-artificial-intelligence-startup-deepmind

Motivation

- Baidu opens Deep Learning Laboratory in Silicon Valley
- Kai Yu from Baidu discusses it.

www.wired.com (April, 2013)



Motivation

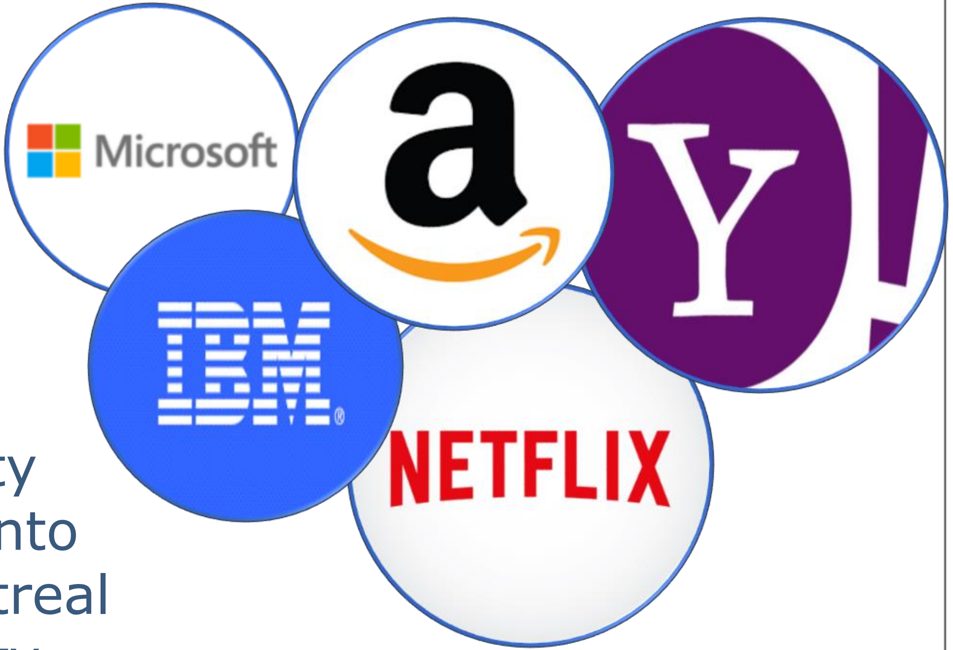
- Baidu hires Andrew Ng
  - Andrew Ng, the man behind Google Brain
  - He led the Google Brain project (a deep learning project)
  - http://www.forbes.com/sites/roberthof/2014/08/28/interview-inside-google-brain-founder-andrew-ngs-plans-to-transform-baidu/

Motivation

Other internet giants using deep learning

- Microsoft
- IBM
- Amazon
- Netflix
- Yahoo

- Universities:
  - Stanford University
  - University of Toronto
  - University of Montreal
  - Newyork University



Strength of Deep Learning

- Deep Learning models have been successful at tasks such as;
- Computer Vision
  - Face detection and recognition.
- Speech Processing
  - Speech Recognition and Speaker Recognition
- Natural Language Processing
  - Machine Translation

Deep Learning

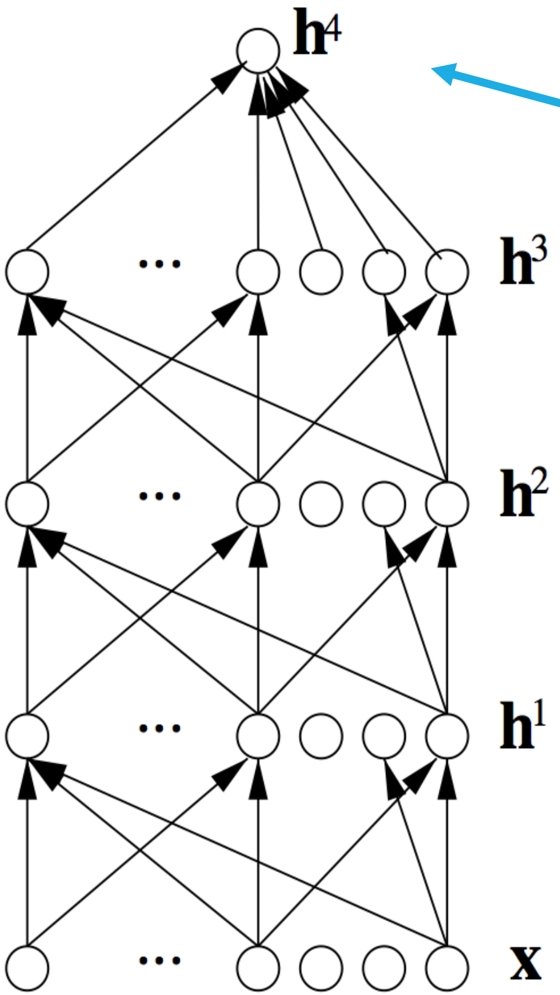
- Deep learning algorithms attempt to learn multiple levels of representation of increasing complexity.
  - With deep learning, Machine Learning becomes just fitting of weights for final decision.

Deep Learning

With Deep Learning, you just give the system a lot of data, so it can discover by itself what some of the concepts in the world are

Prof. Andrew Ng, Standford University
The Man behind the Google Brain
www.wired.com
May, 2013

Deep Architecture



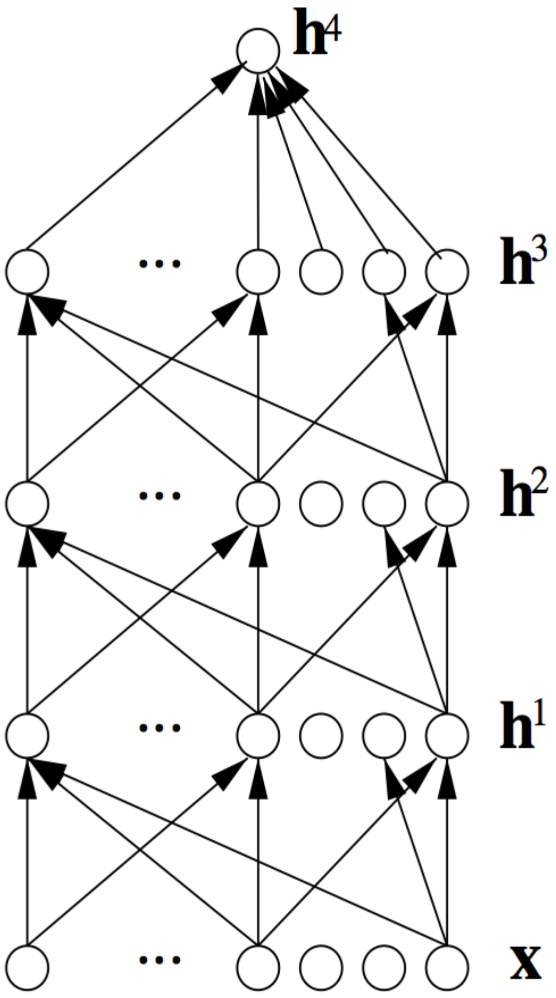
Output Layer
Prediction of the final
target output

Hidden Layers
Learning of more complex
features as we go above

Input Layer
e.g image pixels

Image from Richard Socher

Deep Architecture

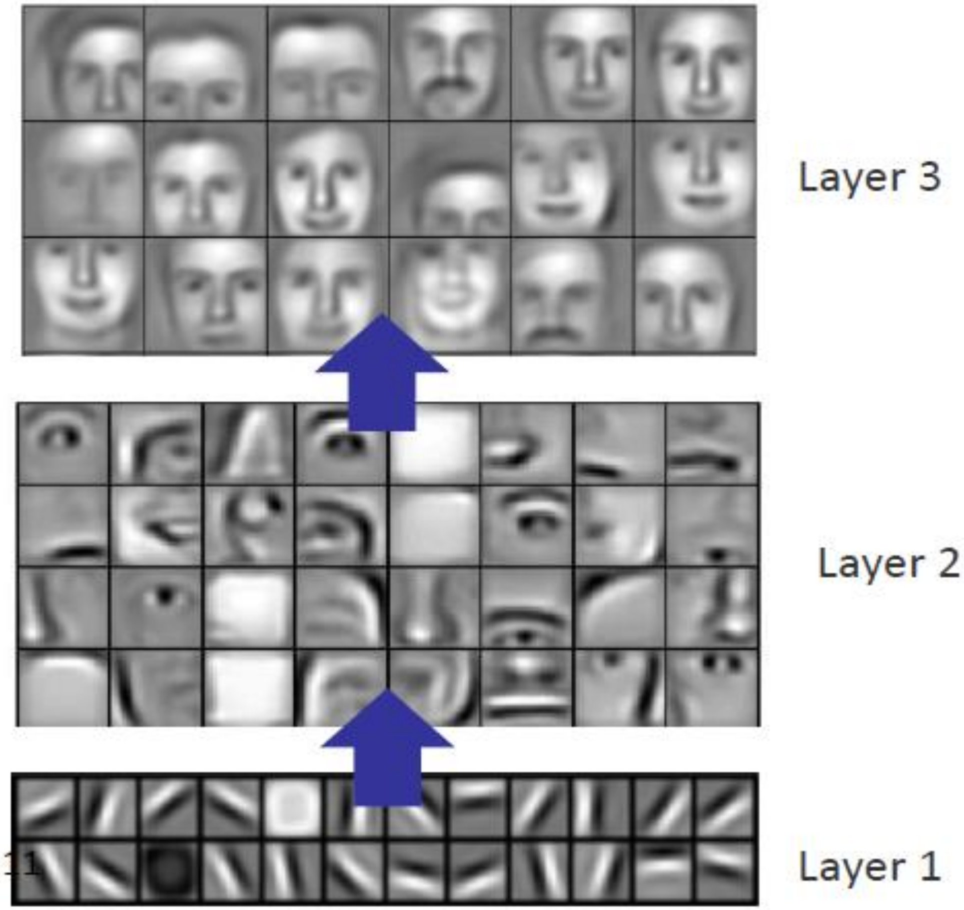


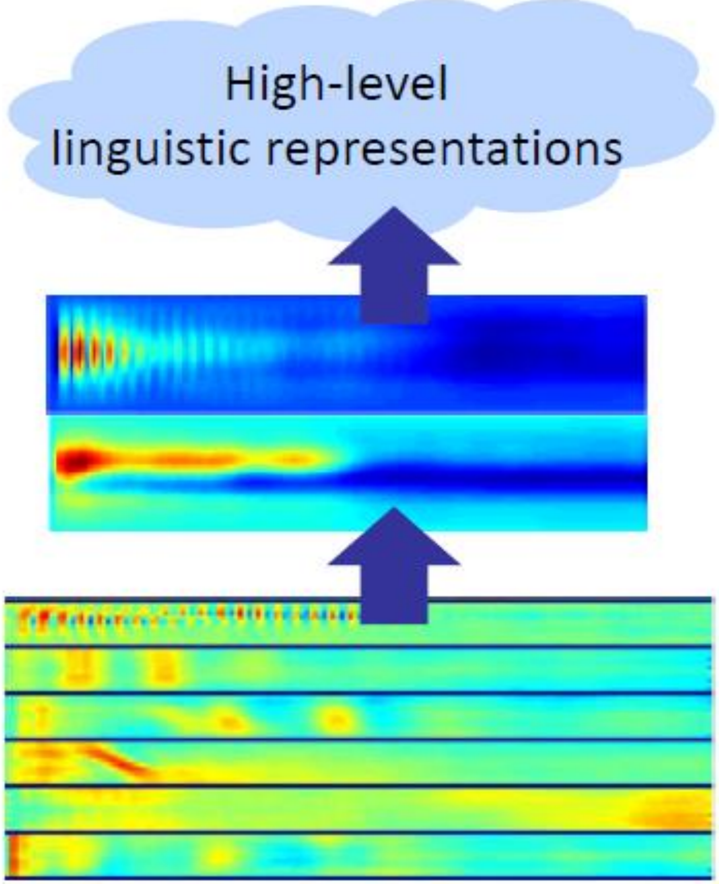
One of the benefit of deep learning is that we can avoid design hand-crafted features.
It is important because, today, most of our data is unlabeled and features learning should be unsupervised.

Image from Richard Socher

Deep Architecture

The higher layers learn more complex and deeper representation.





Lee et al, ICML 2009

Deep Learning

- Beginning?

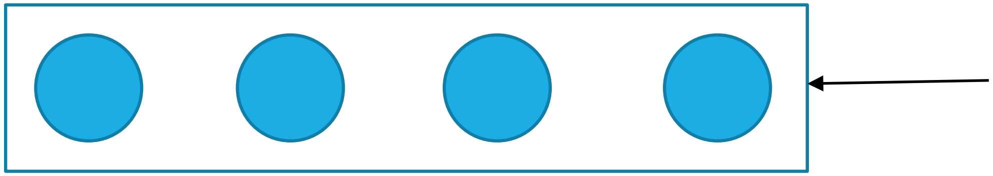
2006

- Efficient algorithms were discovered to train these complex models
- Enough computational resources are available now i.e. faster machines, multi-core CPUs, GPUs.

Break Time

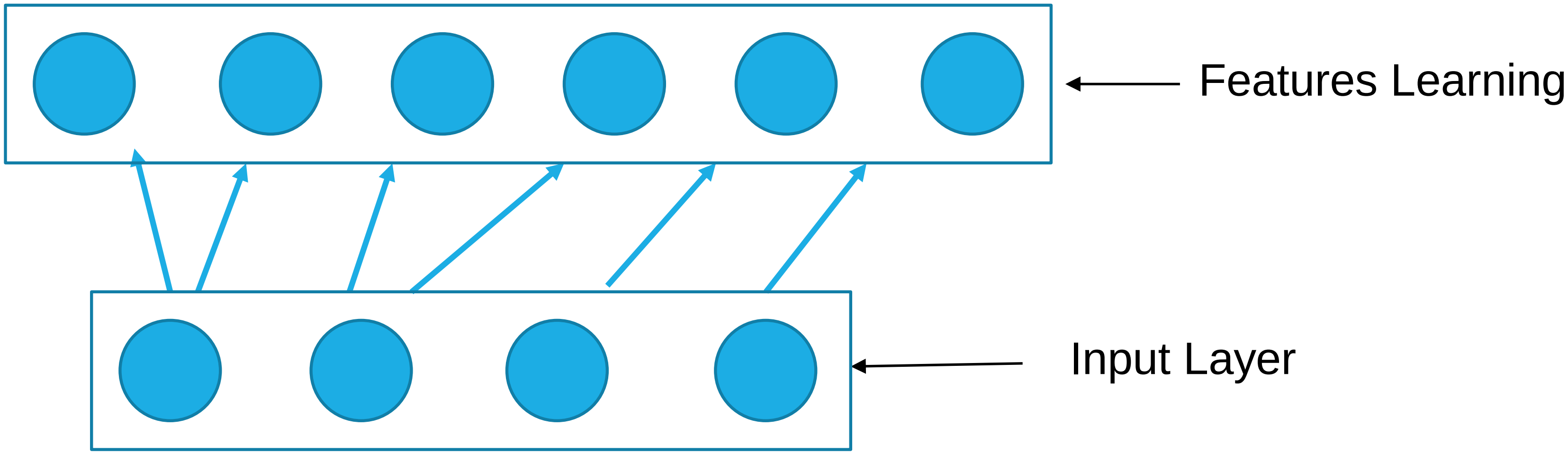
Deep Architecture

Deep Architecture

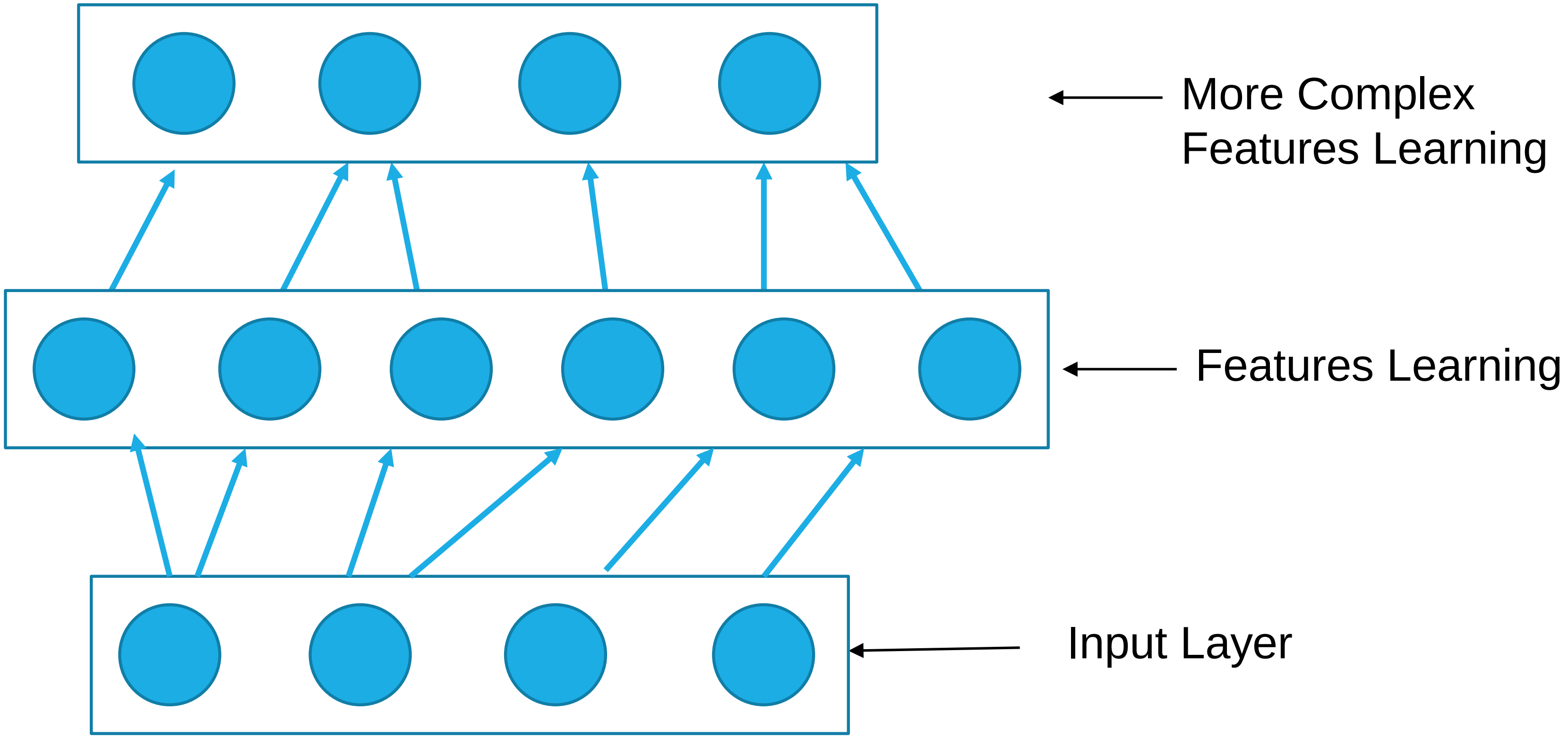


Input Layer

Deep Architecture



Deep Architecture

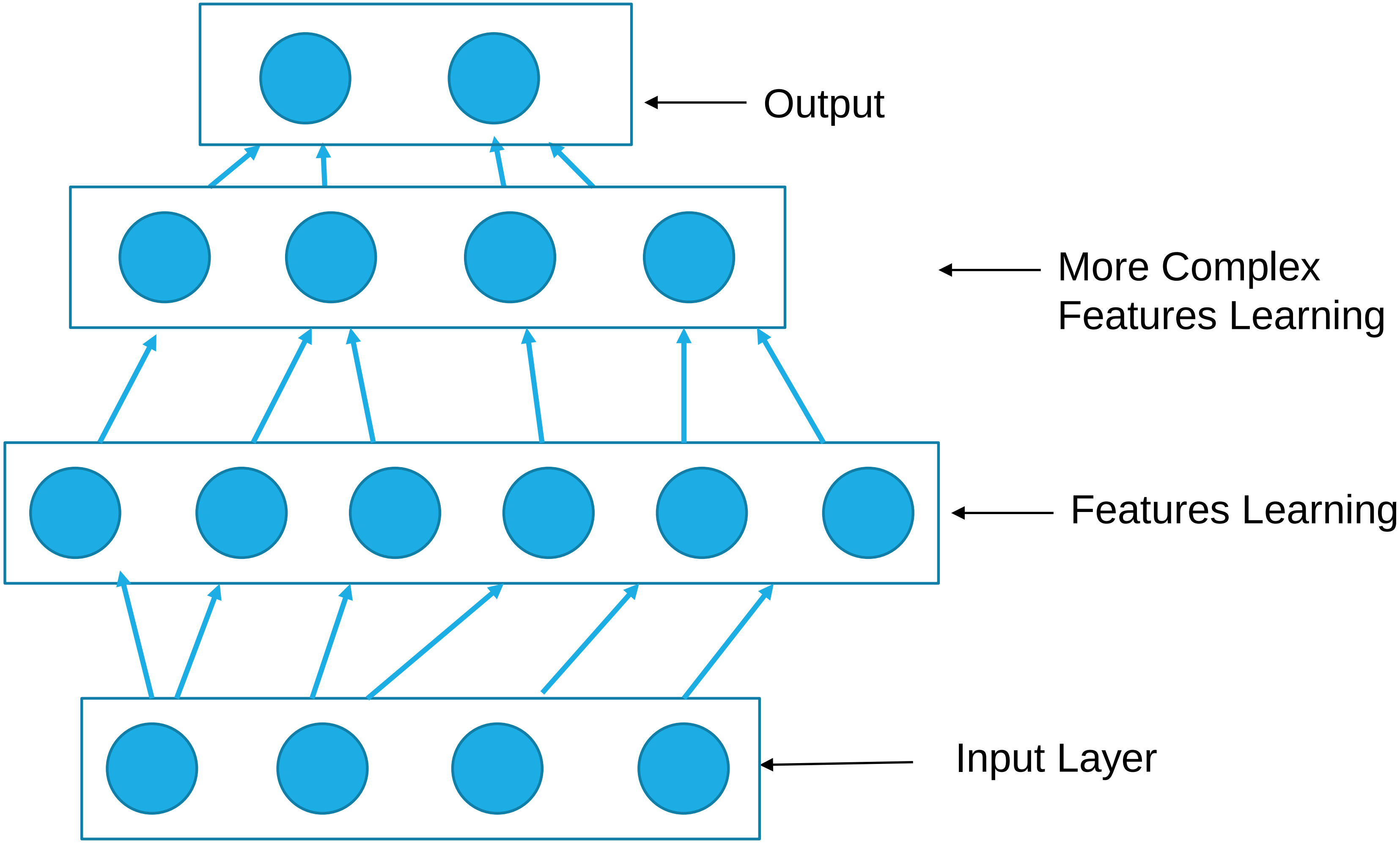


More Complex
Features Learning

Features Learning

Input Layer

Deep Architecture



Output

More Complex
Features Learning

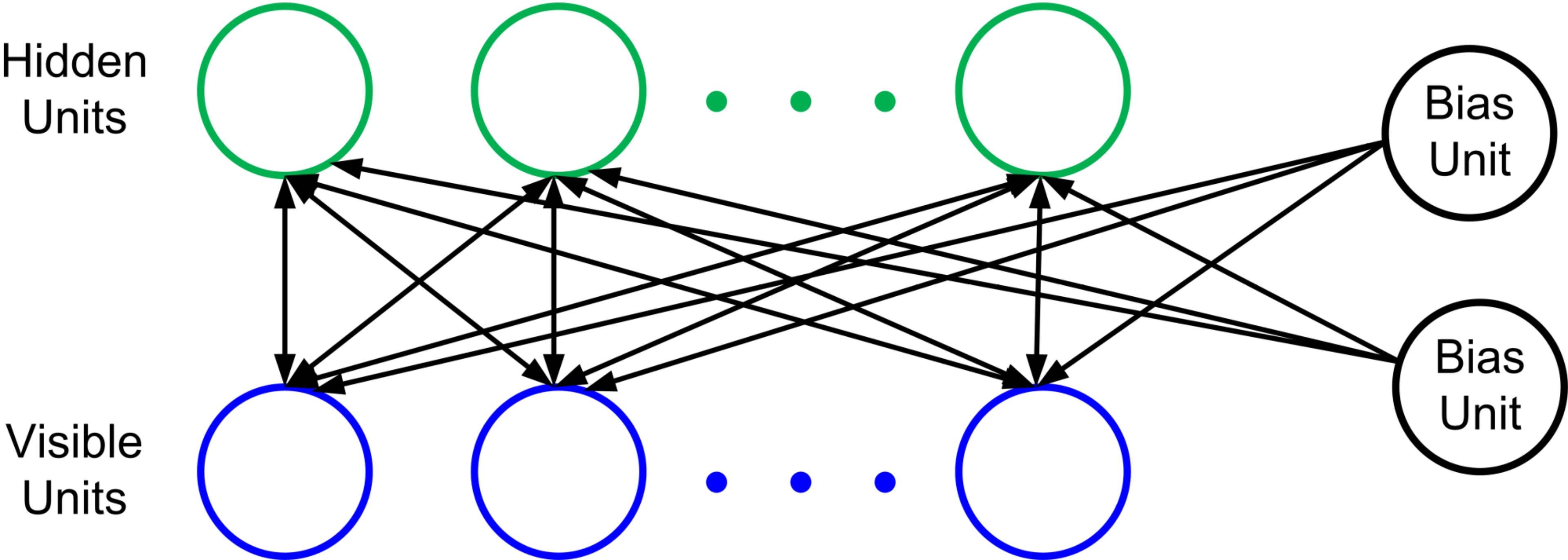
Features Learning

Input Layer

Deep Belief Network

Deep Belief Networks

- The building block of a Deep Belief Network is Restricted Boltzmann Machine (RBM)



Deep Belief Networks

- Restricted Boltzmann Machine
  - Energy based models

$$Energy(x,h) = -b^{'}x - c^{'}h - h^{'}Wx - x^{'}Ux - h^{'}Vh$$

Energy function of Boltzmann Machine

- *W, U,* and *V* are the weight matrices.
- *U* and *V* are symmetric matrices
- *b* and *c* are the bias parameters, associated with *x* and *h* vectors respectively.

Yoshua Bengio: Learning Deep Architectures for AI

Deep Belief Networks

- Boltzmann Machine

$$Energy(x,h) = -b^{'}x - c^{'}h - h^{'}Wx - x^{'}Ux - h^{'}Vh$$

- *For Restricted Boltzmann Machine;*
  - *NO CONNECTIONS BETWEEN HIDDEN-HIDDEN UNITS AND VISIBILE-VISIBLE UNITS*
- Thus, *U=0* and *V=0*

Deep Belief Networks

- Boltzmann Machine

$$Energy(x,h) = -b^{'}x - c^{'}h - h^{'}Wx - x^{'}Ux - h^{'}Vh$$

- Restricted Boltzmann Machine

$$Energy(x,h) = -b^{'}x - c^{'}h - h^{'}Wx$$

- Borrowing equation from Bengio;

$$Free\ Energy(x) = -\beta(x) - \sum_{i} log \sum_{h_i} e^{-\gamma_i(x,h_i)}$$

Putting $\beta(x) = b^{'}x$ and $\gamma(x,h_i) = h_i W_i x$, we get,

$$Free\ Energy(x) = -b^{'}x - \sum_{i} log \sum_{h_i} e^{h_i W_i x}$$

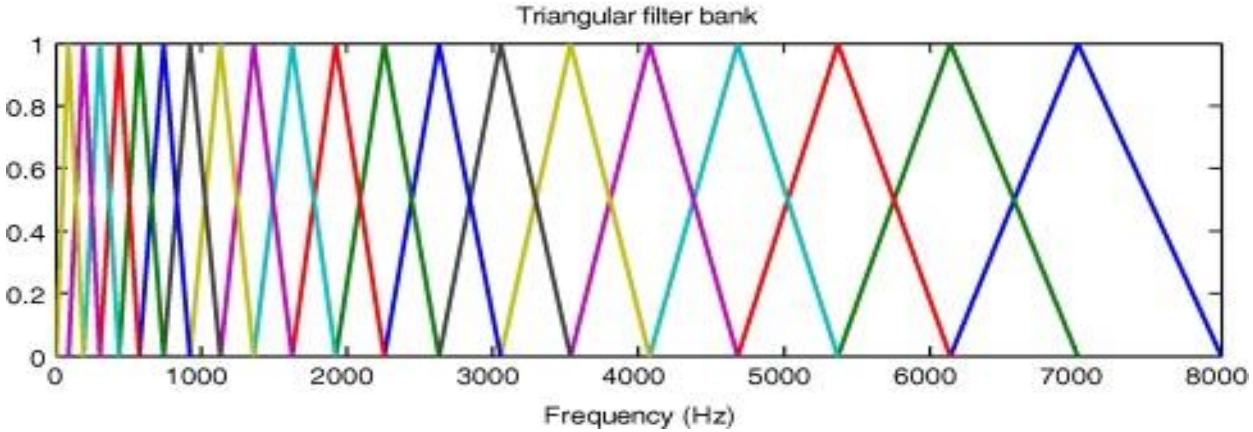
Yoshua Bengio: Learning Deep Architectures for AI

Deep Belief Networks

- The free energy is also referred to be un-normalized log-probability.
- For images, the input units of an RBM are binary.
- However, for speech data, Gaussian inputs units are used (as input is real valued).
- So, the RBM is with Gaussian input units and binary hidden units.

Relevant Technologies

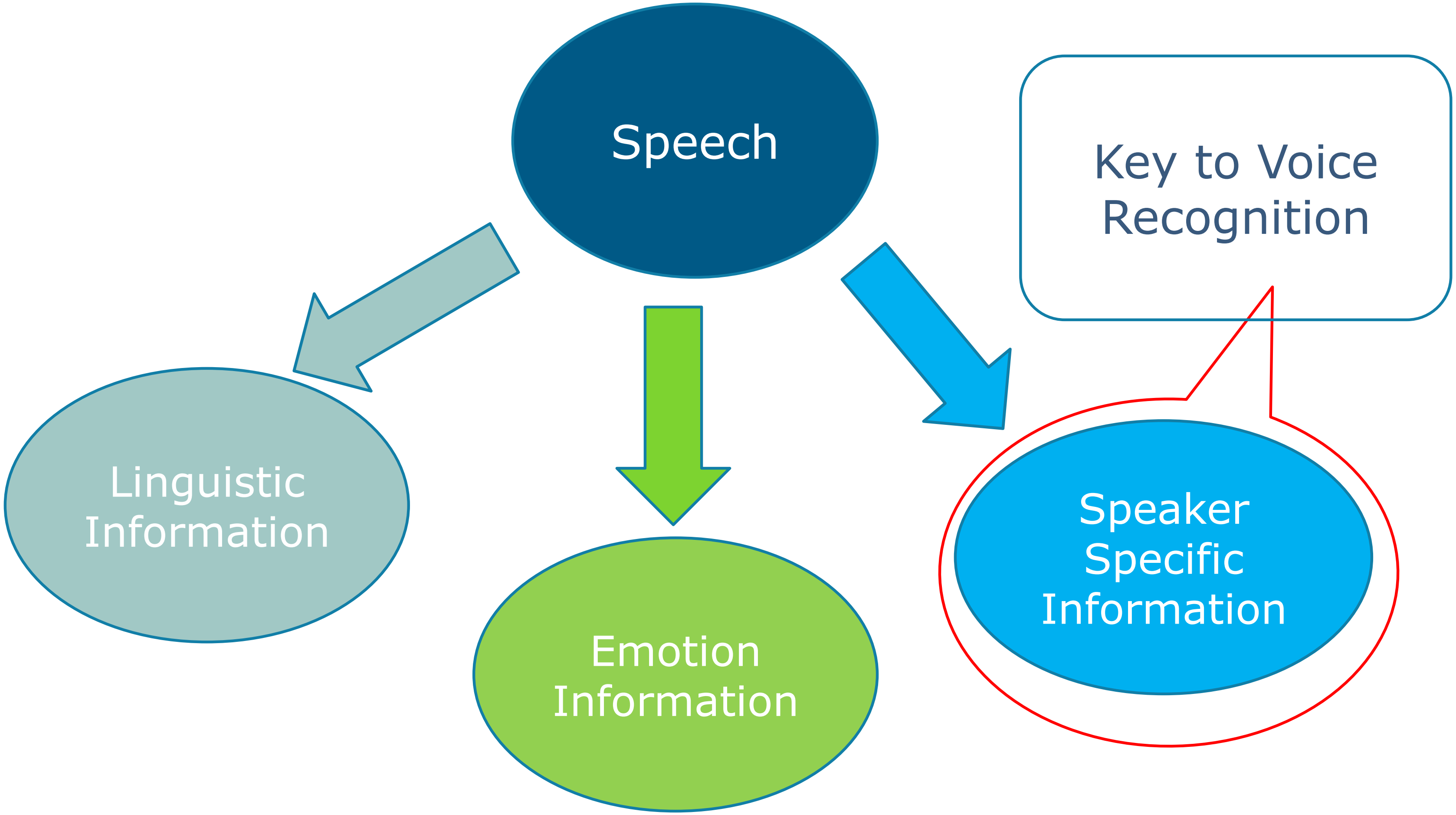
- Mel Frequency Cepstral Coefficients
- Mel-Scale Filter Banks



Mel Scale

- Inspired from Human Ear

Basis for Voice Recognition



Speech

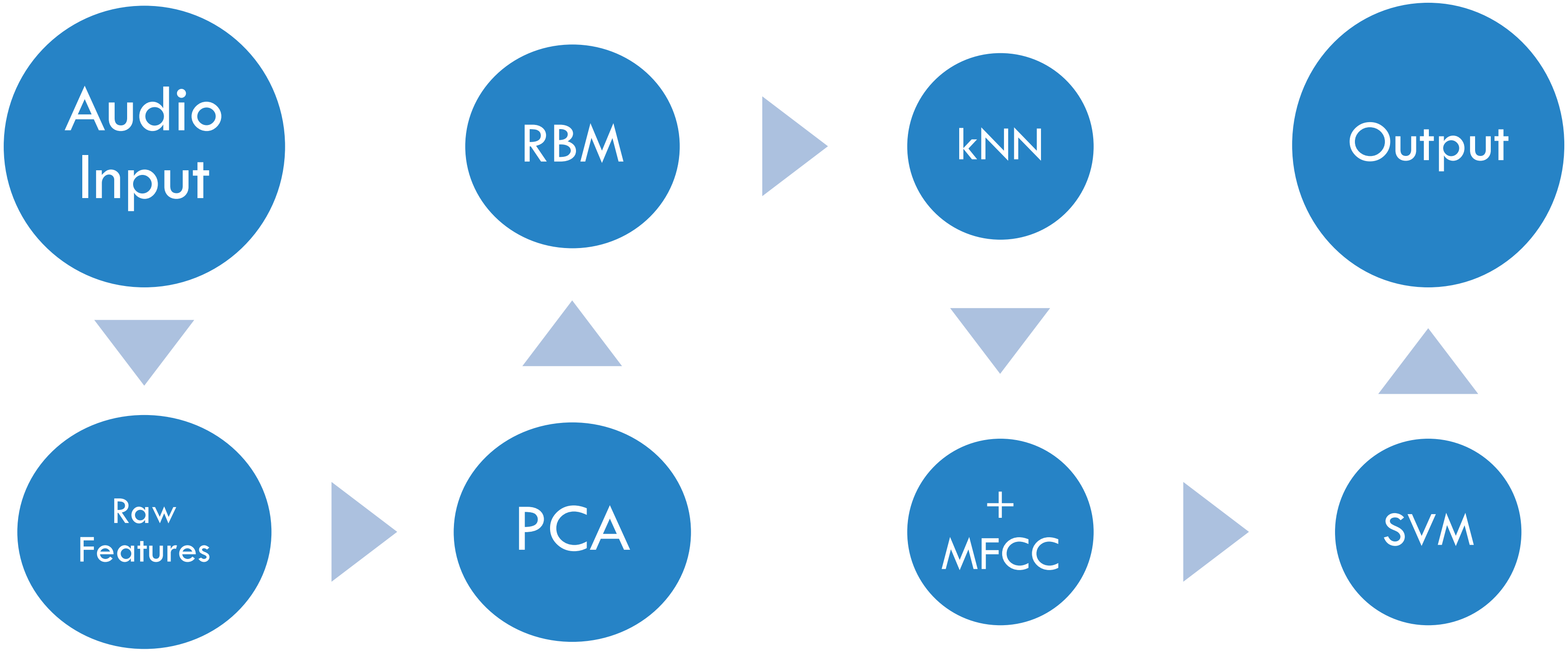
Linguistic Information

Emotion Information

Speaker Specific Information

Key to Voice Recognition

Approach - Demystified



One Example for today

Audio Data Classification

- We combine hand-crafted features with features learnt by RBM and evaluate these combined features for
- Gender Classification
- Speaker Classification

Gender Classification

- RBM Accuracy = 82.5 %
- MFCC Accuracy = 95 %
- RBM + MFCC Accuracy = 97%

Speaker Classification

- RBM Accuracy = 83 %
- MFCC Accuracy = 87 %
- RBM + MFCC Accuracy = 88.88 %

Topics not covered today

- Convolutional Deep Belief Networks
  - Convolutional Neural Networks
- Contrastive Divergence and CD-1
- Recurrent Neural Networks
- Examples of MNIST Digit Recognition
- Number of hidden units
  - E.g 1000 in our network
- Learning rate, momentum
- SVM parameters
- Clustering parameters

Publications

- **H. Ali**, A. S. d'Avila Garcez, S. N. Tran, X. Zhou and K. Iqbal, "Unimodal late fusion for NIST i-vector challenge on speaker detection," *Electron. Lett*., vol. 50, no. 15, pp. 1098–1100, Jul. 2014

- **H. Ali**, N. Ahmad, X. Zhou, K. Iqbal, & S. Muhammad Ali, (2014). DWT features performance analysis for automatic speech recognition of Urdu. *SpringerPlus*, 3(204). doi:10.1186/2193-1801-3-204

- **H. Ali**, X. Zhou, and S. Tie, "Comparison of MFCC and DWT features for automatic speech recognition of Urdu". *In International Conference on Cyberspace Technology (CCT 2013)* pp. 154–158, November 2013, Beijing, China.

Resources

- Deep Learning Tutorials: http://deeplearning.net/tutorials
- Stanford Deep Learning Tutorial http://deeplearning.stanford.edu/wiki/index.php/Main_Page
- Graduate Summer School: Deep Learning, Feature Learning http://www.ipam.ucla.edu/programs/summer-schools/graduate-summer-school-deep-learning-feature-learning/

Conference Proceedings: ICML, NIPS, ICLR etc

Questions