

Issue: **underfitting** due to combinatorially many poor local minima

Culture vs Local Minima

arXiv paper 2012, Y. Bengio

Hypothesis 1

- When the brain of a single biological agent learns, it performs an approximate optimization with respect to some endogenous objective.

Hypothesis 2

- When the brain of a single biological agent learns, it relies on approximate local descent in order to gradually improve itself.

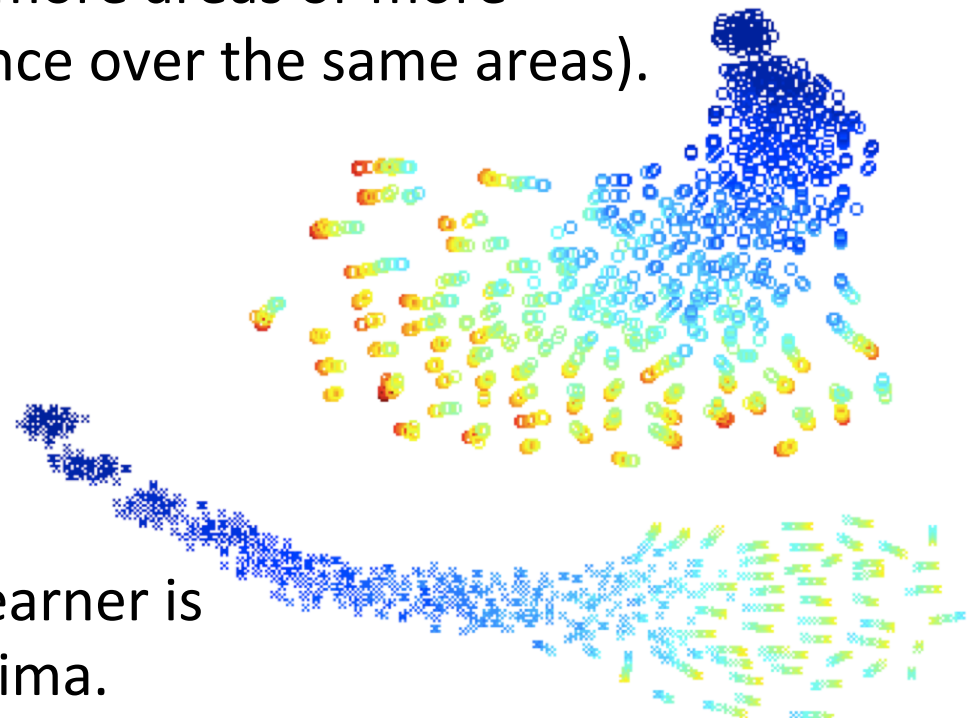
Theoretical and experimental results on deep learning suggest:

Hypothesis 3

- Higher-level abstractions in brains are represented by deeper computations (going through more areas or more computational steps in sequence over the same areas).

Hypothesis 4

- Learning of a single human learner is limited by effective local minima.



Hypothesis 5

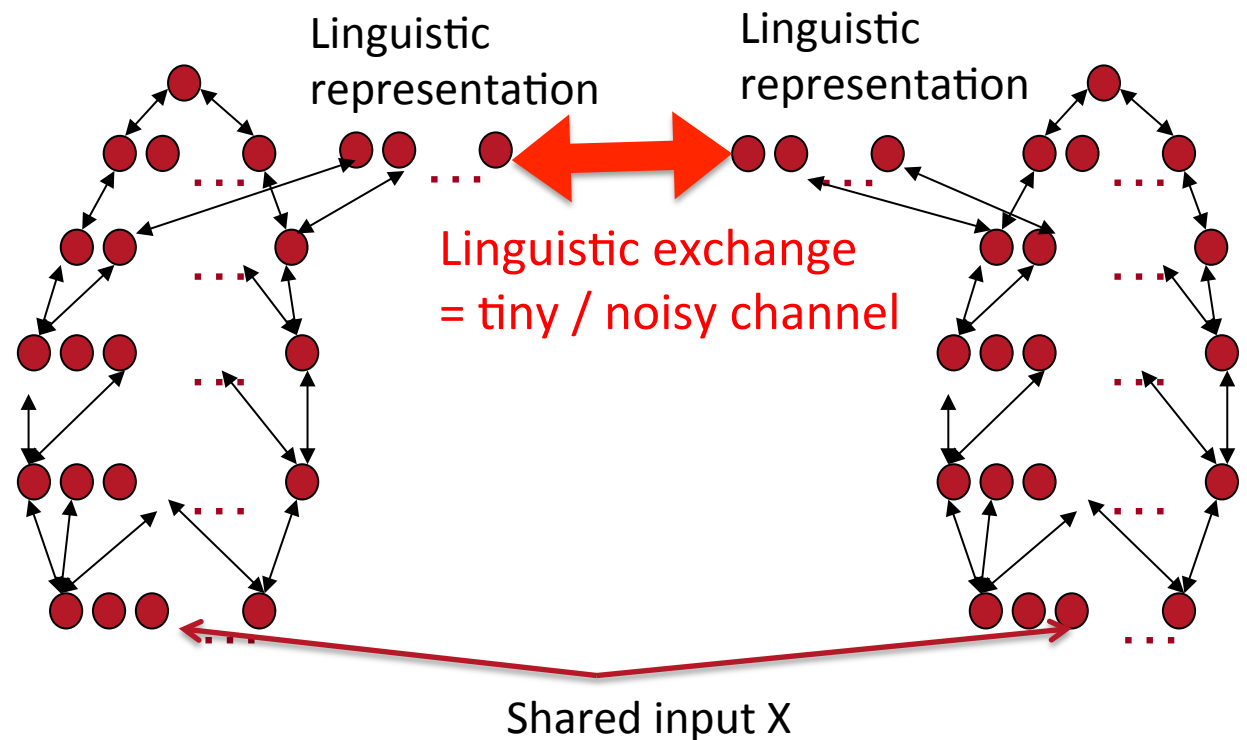
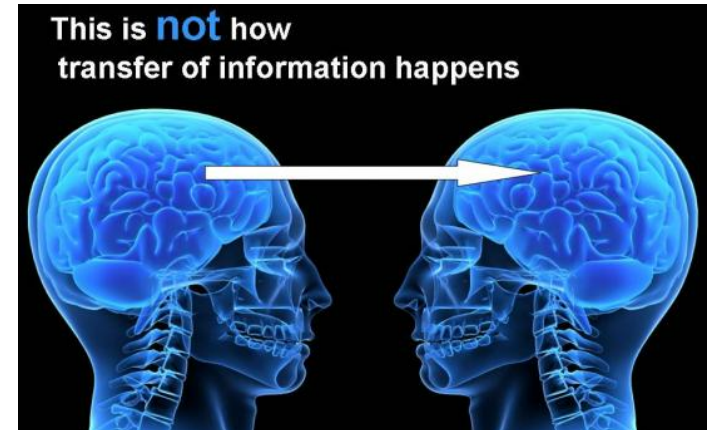
- A single human learner is unlikely to discover high-level abstractions by chance because these are represented by a deep sub-network in the brain.

Hypothesis 6

Curriculum learning

- A human brain can learn high-level abstractions if guided by the signals produced by other humans, which act as hints or indirect supervision for these high-level abstractions.

How is one brain transferring abstractions to another brain?



How do we escape local minima?

- linguistic inputs = extra examples, summarize knowledge
- criterion landscape easier to optimize (e.g. curriculum learning)
- turn difficult unsupervised learning into easy supervised learning of intermediate abstractions

How could language/education/culture possibly help find the better local minima associated with more useful abstractions?

Hypothesis 7

More than random search:
potential exponential speed-up by divide-and-conquer
combinatorial advantage:
can combine solutions to
independently solved sub-problems

- Language and meme recombination provide an efficient **evolutionary operator**, allowing rapid search in the space of **memes**, that helps humans build up better high-level internal representations of their world.

From where do new ideas emerge?

- Seconds: **inference** (novel explanations for current x)
- Minutes, hours: **learning** (local descent, like current DL)
- Years, centuries: **cultural evolution** (global optimization, recombination of ideas from other humans)

Related Tutorials

- See “*Neural Net Language Models*” **Scholarpedia** entry
- Deep Learning tutorials: <http://deeplearning.net/tutorials>
- Stanford deep learning tutorials with simple programming assignments and reading list
<http://deeplearning.stanford.edu/wiki/>
- ACL 2012 Deep Learning for NLP tutorial
<http://www.socher.org/index.php/DeepLearningTutorial/>
- ICML 2012 Representation Learning tutorial <http://www.iro.umontreal.ca/~bengioy/talks/deep-learning-tutorial-2012.html>
- **More reading: Paper references in separate pdf, on my web page**

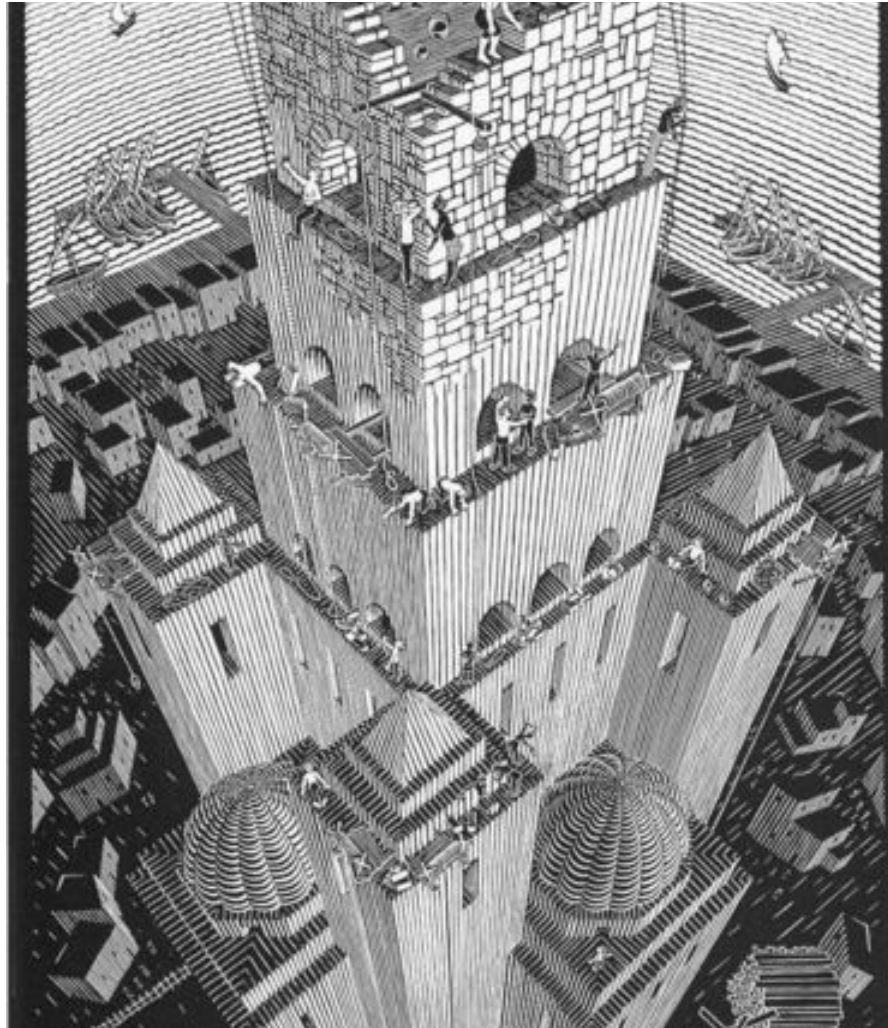
Software

- Theano (Python CPU/GPU) mathematical and deep learning library <http://deeplearning.net/software/theano>
 - Can do automatic, symbolic differentiation
- Senna: POS, Chunking, NER, SRL
 - by Collobert et al. <http://ronan.collobert.com/senna/>
 - State-of-the-art performance on many tasks
 - 3500 lines of C, extremely fast and using very little memory
- Recurrent Neural Network Language Model
<http://www.fit.vutbr.cz/~imikolov/rnnlm/>
- Recursive Neural Net and RAE models for paraphrase detection, sentiment analysis, relation classification www.socher.org

Software: what's next

- Off-the-shelf SVM packages are useful to researchers from a wide variety of fields (no need to understand RKHS).
- To make deep learning more accessible: release off-the-shelf learning packages that handle hyper-parameter optimization, exploiting multi-core or cluster at disposal of user.

The End



LISA team: **Merci! Questions?**

