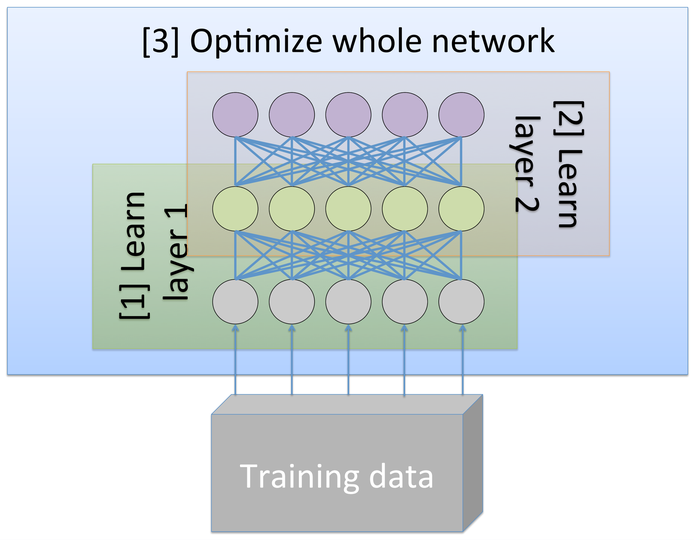
Progressive Development of the Number Sense in a Deep Network

Will Y. Zou, James L. McClelland

1 Abstract

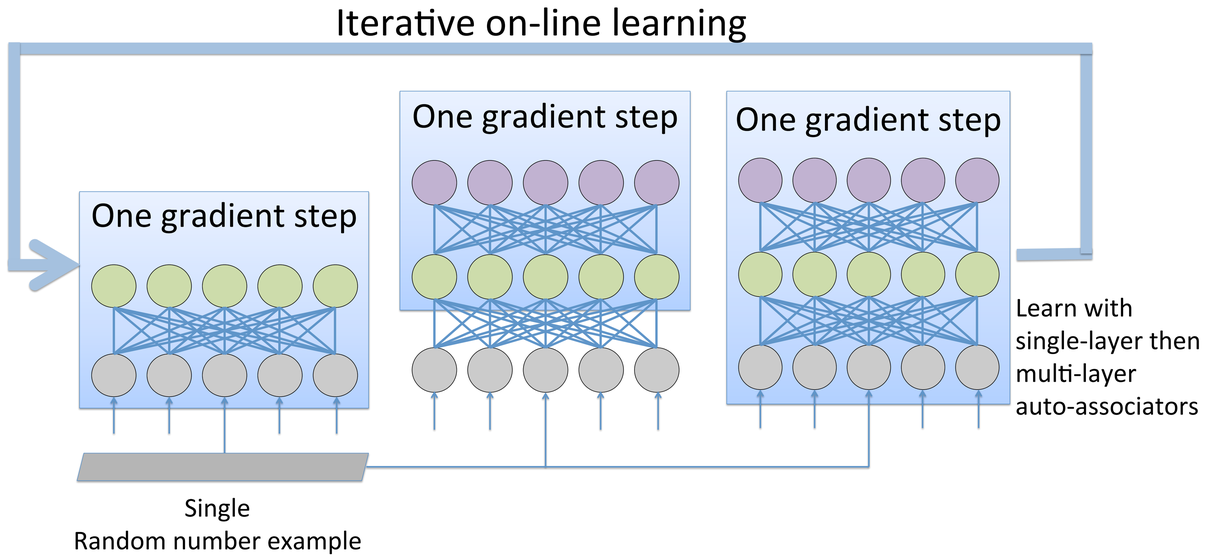
What are the developmental bases of the number sense? This ability could arise through evolution or experience. Stoianov & Zorzi [1] showed that a neural network could learn number sense from visual examples containing varying numbers of elements. However, the layer-wise training regime is unrealistic from a developmental standpoint. A key observation is that number acuity progressively develops from infancy to adulthood (as reflected by a decreasing Weber fraction). This development involves accumulation of single examples, each of which updates the connection weights in a hierarchical system. We present an unsupervised deep network that learns all weights as it observes one `number example' at a time. As on-line training progresses, neurons representing numerosity start to emerge in the deeper layers, and the Weber fraction progressively sharpens. These results establish that a generic learning algorithm in a deep network gives rise to a clear developmental trajectory of the number sense.

2 Progressive Learning Algorithm for Cognitive Development



Sample training data
(each black square is one example)

```
Algorithm
Iteratively draw a random
number example:
    - learn first layer
      with auto-association
    - learn second layer
      with auto-association
      (with first layer fixed)
    - learn both layers
```



DBN [2] layer-wise learning:
Developmentally infeasible

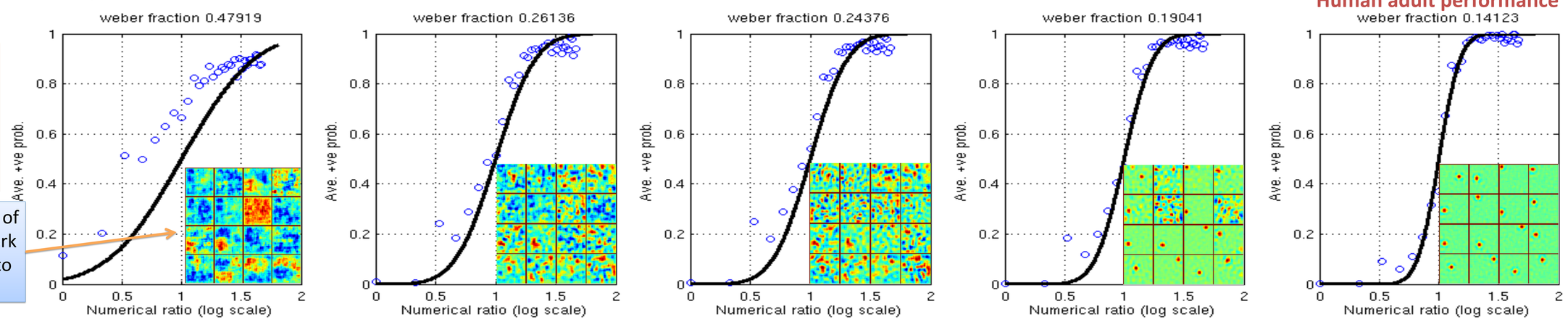
Progressive Stochastic Gradient Descent:
Simulates cognitive development by observing one `number' example at a time

3 Progressive Development of the `Number Sense'

After unsupervised learning, the weber fraction for number discrimination is determined with second layer features.

The neural receptive field of the artificial neural network progressively develops into center-surround filters.

Neurons are developed on the top which are (Red Circles) number selective invariant to cumulative area. (Green Circles) cumulative area selective invariant to number. (Blue Pluses) Other neurons which explain significant variance in the data (Black Crosses) Non-significant neurons



Progressive Development in the Deep Neural Network
with respect to # iterations

Regression experiment for finding number neurons

4 Experimental Details

Neural Network Architecture: we use 80 units on the first layer, 400 units in the second layer, and apply the sigmoid non-linearity to each hidden unit.

Weber fraction: a linear classifier is trained on the second layer features to determine whether a number is larger or smaller than eight. An error function, whose variance is proportional to the Weber Fraction, is fitted to the scatter plot of average classifier probability versus log ratio of numbers.

Regression experiment for finding number neurons: we use weighted regression to find the solution (a, b) to $h_j = a \log(N) + b \log(A) + e$, for each neuron j, and plot on the two dimensional space. The `number neurons' are selected by making sure that the regression explains at least 10% of the variance in the data.

5 Conclusion

We present the Progressive Stochastic Gradient Descent algorithm for on-line learning in deep neural networks:

- A feasible algorithm for modeling progressive cognitive development for deep neural networks
- Learns different layers of the deep neural networks iteratively and reaches reasonable local minima
- Updates progressively by observing one stochastically chosen data example at a time

Using this developmentally feasible algorithm, we obtain results resembling those obtained in human cognitive experiments:

- Developmental progression of sharpening Weber Fraction in visual number discrimination
- Number-sensitive neurons invariant to cumulative area emerge on the top layer of the deep network

6 References and Acknowledgement

[1] I. Stoianov and M. Zorzi, Emergence of a 'visual number sense' in hierarchical generative models. *Nature Neuroscience, 2012,* 8, 194-196

[2] G. E. Hinton and R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks. *Science. Vol. 313. no. 5786, pp. 504 - 507, 28 July, 2006*

J. L. M was supported by AFOSR. We thank Andrew Saxe and Adam Coates for helpful discussions.