Deep learning from scratch: exploring generalization performance in digit classification

Sofía Lawrie
sofia.lawrie@upf.edu
Department of Information and Communication Technologies and Center for Brain and Cognition
Universitat Pompeu Fabra, Barcelona, Spain

INTRODUCTION

Humans have an outstanding generalization power for visual categorization: we can recognize multiple instances of a class of objects even after seeing only one exemplar. This property constitutes the Holy Grail for researchers in the computer vision field, but also draws the attention of researchers in the AI community in general, as well as cognitive neuroscientists and psychologists, due to the modern mutualism between these areas.

In the past few years, the rise of deep learning has boosted the efforts to create the ultimate artificial vision system, with record-breaking performances on complex datasets such as ImageNet. However, these rich algorithms often require large amounts of high-quality data, with many variations of objects belonging to the same class, to be trained on. In the present work, we make a short demonstration of how a simple deep learning architecture can achieve good performance in a digit classification task, but also show how it cannot learn features that remain invariant to affine transformations of the digits. On the other hand, when exemplars belonging to the affine-transformed dataset are used for training, the network improves its performance significantly on the transformed test set.

NEURAL NETWORK ARCHITECTURE

The neural network implemented contained two hidden layers (500 and 150 units each), as well as input and output layers (Figure 1). All of the units had sigmoid non-linearities. Backpropagation and a mini-batch gradient descent implementation were employed to minimize the cross- entropy cost function with an L2- regularization term. All the network hyper-parameters were tuned using grid-search on Training Set 1, defined in Datasets.

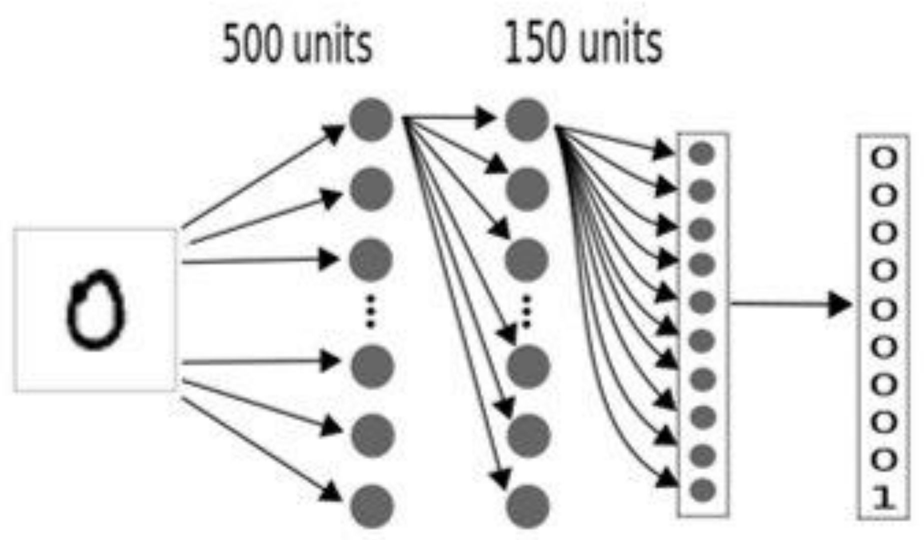

Figure 1. Schematic illustration of neural network used. Input layer is represented by the digit. Only connections from the first unit in each hidden layer are shown. Bias units are omitted.

DATASETS

The datasets used belong to affNIST [1] . Training Set 1 corresponds to the just-centered version of affNIST training-validation dataset (the original 60,000 28x28 images from MNIST [2] remapped to 40x40 images). Test Set 1 corresponds to 10,000 images from the just-centered version of affNIST test set (Figure 2). Training Set 2 corresponds to all 60,000 images from a batch of the transformed version of the affNIST training and validation dataset. Test Set 2 corresponds to 10,000 randomly sampled images from the transformed version of the affNIST test set (Figure 3).

Figure 2. Random samples from the just-centered version of affNIST.

Figure 3. Random samples from the transformed version of affNIST.

RESULTS

For each training set, ten different instances of the network were initialized and trained. Mean performance and sem are reported for each test set on Table 1.

Classification accuracy		
Training Set	Test Set 1	Test Set 2
1	0.9758(3)	0.137(1)
2	0.898(2)	0.816(1)

Table 1. Mean (sem) network performances across ten instances.

DISCUSSION

- The long-known problem of learning mappings between inputs and outputs that remain invariant to certain transformations is easily observed in the results from Table 1, as there is a large drop in accuracy when testing the network in digits that were affine -transformed when compared to digits that appeared centered within the image frame.
- If we consider that the full input space is composed by all the possible instances of every digit, poor generalization is not surprising, as the network is only exposed to a very small sub-space of this distribution and tunes its parameters to describe it as accurately as possible, while the transformed digits present a bigger amount of intra-class variability, which remains as a largely unexplored region of input space (Figure 4). Thus, *locality* [3], a central assumption of many machine learning algorithms, is violated and performance drops.

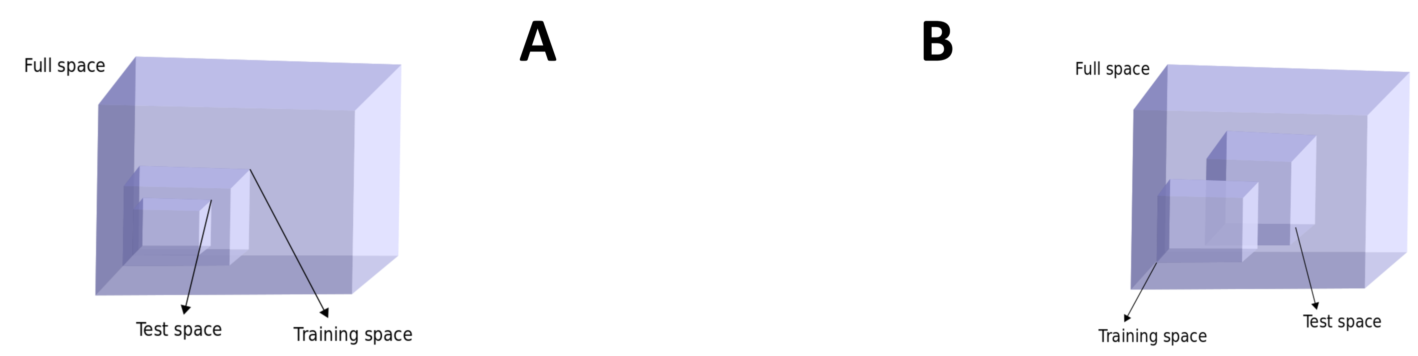

Figure 4. A. Schema of probability space seen by the network in Training Set 1 and Test Set 1. As locality is not violated, the network has better performance. B. Schema of probability space seen by the network in Training Set 1 and Test Set 2. As there is little overlap between the spaces and the locality assumption does not hold, performance drops.

- The drop in accuracy in Test Set 1 when using Training Set 2 is mostly explained by the fact that, during training, the network sees comparatively less examples from the region of input space it is subsequently tested on. However, considering that the accuracy is still better for Test Set 1 than Test Set 2 indicates that the network makes computations that are biased towards learning mappings that better represent the easy centered cases. This could potentially also be boosted by the hyper-parameters selection, as they were tuned for Training Set 1.
- Lacking good and representative datasets of the input distributions can have large negative effects on naive neural networks.

CONCLUSIONS

- We have shown that naïve feed-forward neural networks cannot make inferences in regions of input space they have not properly explored during training.
- The performance of a classification algorithm depends on architecture style and optimization, but most of all, on the quality of the dataset used for training.
- It is important to design architectures that build desired properties into the network, when these are difficult to directly learn from the dataset. Such is the case for Convolutional Neural Networks, whose design introduces translation invariance into the learned mappings [3].
- The simple example shown here illustrates the need for the problems of Transfer Learning [4] or Domain Adaptation [5] to be properly addressed, as they could help improve algorithms performance in situations where the data probability space is not extensively sampled.

REFERENCES

[1] https://www.cs.toronto.edu/~tijmen/affNIST/
[2] http://yann.lecun.com/exdb/mnist/
[3] C.M. Bishop. Pattern Recognition and Machine Learning. 2006, Springer-Verlag New York, Inc., Secaucus, NJ, USA
[4] R. Sousa, L. M. Silva, L. A. Alexandre, J. Santos and J. Marques de Sá. Transfer Learning: Current Status, Trends and Challenges. 20th Portuguese Conference on Pattern Recognition, 2014.
[5] G. Csurka. Domain Adaptation for Visual Applications: A Comprehensive Survey. arXiv:1702.05374

S.L is gratefully funded by a UPF-PhD scholarship from the Government of Catalonia and would also like to thank Rubén Moreno-Bote for proof-reading and extra-funding.