A Sample-and-Learn scheme for Neural Networks

Andreas Antoniades, Clive Cheong Took
a.antoniades@surrey.ac.uk, c.cheongtook@surrey.ac.uk

UNIVERSITY OF SURREY

Objectives

- Reduce neural network training time.
- Maintain acceptable accuracy levels.
- Render neural networks scalable to big data.

Introduction

Humans spend the entirety of their lives learning. Just like the human brain, neural networks are plagued by prolonged training times and this work aims to alleviate that issue.

Sample-and-Learn Method

▷ Learning is performed at a small scale.

√ Reservoir sampling samples the dataset once per epoch.

▷ Provide exposure to a large scope of the data.

√ Data is explored through the epochs.

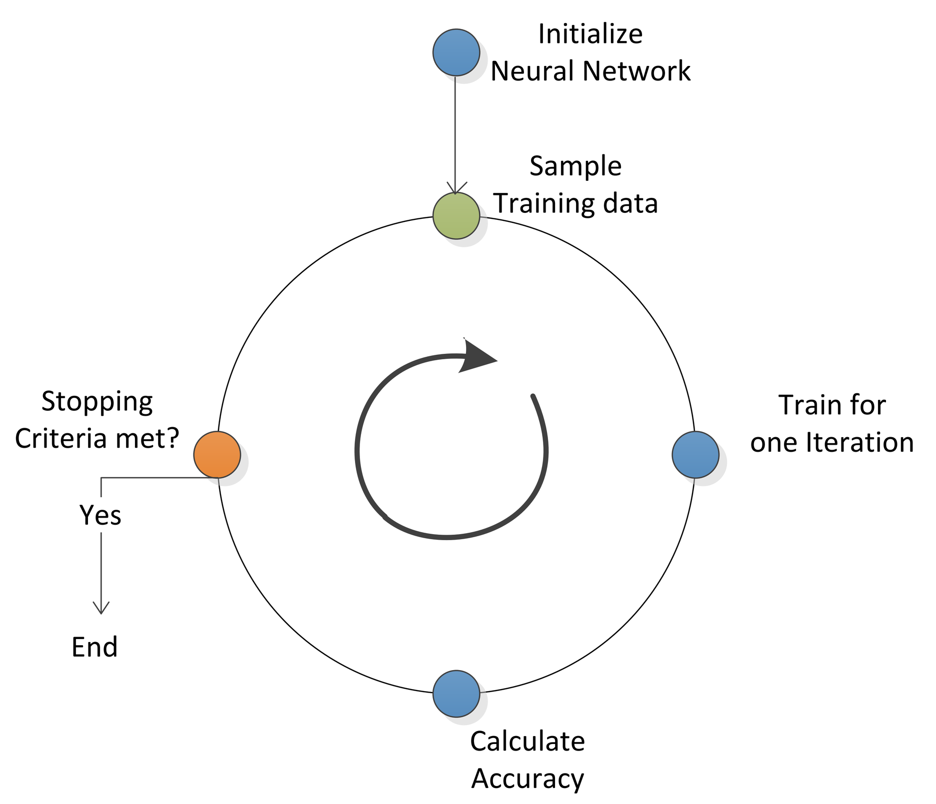

Figure 1: Diagram of the proposed sampling-based training approach.

Conclusion

We demonstrated a significant reduction in computational cost when a neural network learns from smaller randomly sampled data rather than learning from the entire data set.

By applying our Sample-and-Learn scheme, the training time of the neural network was reduced by an order of magnitude without affecting the accuracy.

References

- G. Hebrail, UCI Machine Learning Repository, https://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption. EDF R&D, France, 2010.
- A. Antoniades, C. Cheong Took, "When small data helps big data learning", submitted to Elsevier Neural Networks, Special Issue on Neural Network Learning in Big Data, Feb 2015
- A. Antoniades, C. Cheong Took, "A Sampling-based Neural Network classifier", submitted to International Joint Conference on Neural Networks, 2015.

A Neural Network Simulation

Factors that affect accuracy:

- Size of data set (N)
- Epochs for training (E)
- Size of training data per epoch (s)

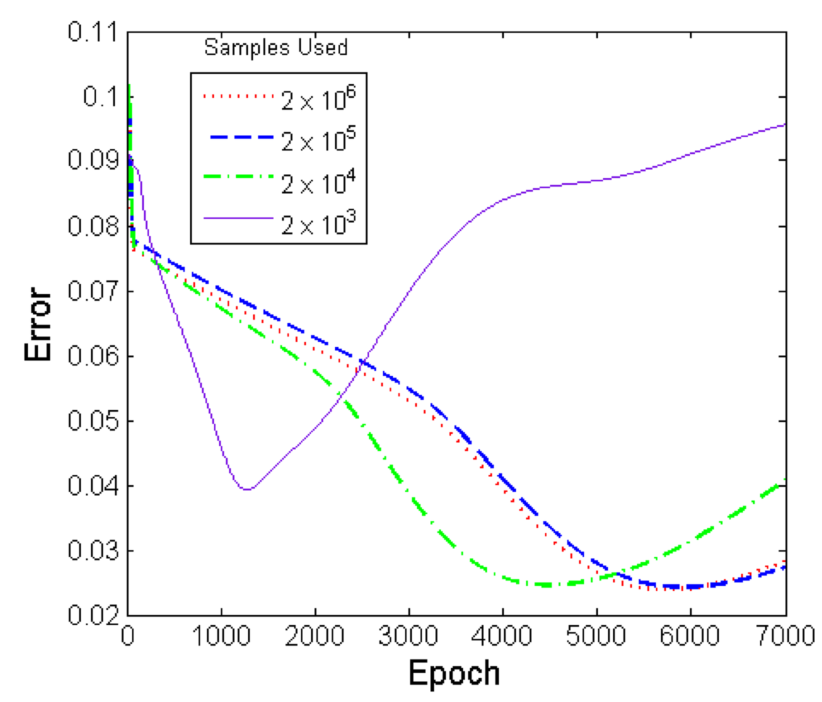

Figure 2: Learning curves for test error for different datasizes.

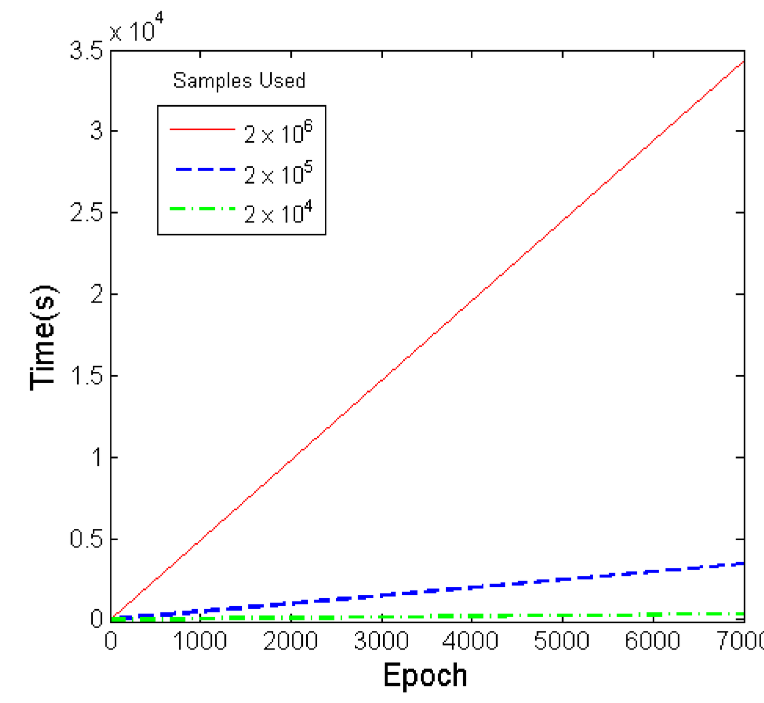

Figure 3: Running time comparison for different datasizes.

Sample-and-Learn Method - Analysis

Probability that a sample k is in the reservoir after one-pass through the entire data set $\rightarrow \frac{s}{N}$
The probability of getting kth sample at least once through E epochs is:

$$p(k)' = 1 - \left[1 - \frac{s}{N}\right]^E \tag{1}$$

E or s which one affects time complexity the **most**?
Rate of change of $p(k)'$ with respect to both s and E is derived as follows:
Let $X = 1 - \frac{s}{N}$

$$\frac{\partial p(k)'}{\partial E} = -X^E \ln(X) \tag{2}$$

$$\frac{\partial p(k)'}{\partial s} = \frac{E}{N}\left[X\right]^{E-1} \tag{3}$$

For clarity, it is also instructive to determine the relationship between Δs and ΔE

$$\frac{\partial p(k)'}{\partial E} = -\frac{N}{E}\frac{\partial p(k)'}{\partial s} X \ln(X) \tag{4}$$

X in (4) is a fraction, hence $ln(X)$ is a negative number. By letting $\ln(X) = -\lambda$, Equation (4) becomes

$$\frac{\partial p(k)'}{\partial E} = \lambda X \frac{N}{E}\frac{\partial p(k)'}{\partial s} \tag{5}$$

Using the approximation $\partial y/\partial x \approx \Delta y/\Delta x$, where Δ is the change in variable x or y and taking the inverse of (5), the relationship between a change in number of epochs ΔE and a change in data size Δs can be approximated as

$$\Delta E \approx \frac{E}{\lambda X N}\Delta s \tag{6}$$

Remark: To achieve the same level of accuracy, training with a smaller data size does not cause a significant increase in the number of epochs.

Experimental Results

Experimental results of our Sample-and-Learn algorithm with a feed-forward neural network. We have used the individual household electric power consumption data set for our simulations.

Data size	Test Error	Accuracy	Epochs	Time(s)
20652596	0.0167	98.3%	5897	27532 ⟸ Full Dataset without Sample-and-Learn
100,000	0.0167	98.3%	4910	21832
50,000	0.0174	98.2%	4907	5657
10,000	0.0168	98.3%	7489	1029
7,000	0.0166	98.3%	12526	1251
3,000	0.0178	98.1%	26138	1369
1,000	0.0169	98.3%	31246	1385
500	0.0169	98.2%	37041	1203
100	0.0167	98.3%	39013	1059

Table 1: The computational cost and accuracy of the Sample-and-Learn Scheme over different reservoir sizes