MINST Kaggle Digit Recognizer: Contrasting the Randomized Forest andMLP Neural Network

Andrew Osborne
amo004@uark.edu

Josh Price
jdp024@uark.edu

April Walker
adw027@uark.edu

University of Arkansas,
Fayetteville, AR, 72701, USA

Abstract

In order to tackle the classic MINST dataset, our team developed a randomized forest and a multi-layer perception (MLP) classifier using the sklearn Python library. While our RF was able to achieve about 6% test error with only 10 trees, increasing the number of trees soon developed diminishing returns. Cross validation never breached 97% accuracy on test data, suggesting an RF is not well suited to solve such a high dimensional problem. In order to optimize our MLP NN, we explored the use of stochastic and Adam gradient descent, however found the later consistently outperformed the former. In order to balance accuracy vs. complexity, our cross validation method suggests the use of 1 hidden layer and 256 nodes. Our final Kaggle submission resulted in a correct prediction rate of 97.90%, suggesting our NN is well-trained.

1 The MINST Dataset

I'll probably talk about the considerations going into this project and the reason behind our chosen models. It might also be nice to just give a general background of the dataset that will set this up?

2 Randomized Forest

Here I will give a basic whatever on RF's that Dr. Zhang will already know. "It's like a decision tree with extra steps that gives you 'good enough' results." -Someone

2.1 Implementation

This is probably not the best title for this section

2.2 Cross-Validation and Results

2.3 Future Considerations

Because an RF is probably so so, but there are probably similar models that work better on problems w/ high levels of dimensionality.

3 Multi-Layer Perceptron Classifier

This will be much more interesting to talk about

3.1 Contrasting Gradient Descent Algorithms

Because it will probably be good to talk about why Adam's so good, basically because it used the components of a lot of other good methods.

3.2 Cross-Validation and Results

4 Conclusions

MLP NN's are noice, Los Bobo's is better.

References

References follow the acknowledgments. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. **Remember that you can use more than eight pages as long as the additional pages contain *only* cited references.**

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.