An Empirical Evaluation of Sketched SVD and its Application to Leverage Score Ordering: Supplementary Material

Author Name1　　　　　　　　　　　　　　　　　　　　　　　abc@sample.com
Address 1

Author Name2　　　　　　　　　　　　　　　　　　　　　　　xyz@sample.com
Address 2

1. Model Configurations

Here we provide more details on the model configurations used to ensure reproducibility of results.

1.1. Models for MNIST dataset

For the MNIST dataset, we implement the following models:

LR is a Logistic Regression classifier (Collins et al., 2002). **NN Small** is a small non-linear neural network (Haykin, 1998) with a hidden layer of 784 units with ReLU activation. **NN Large** is a large non-linear neural network (Haykin, 1998) with 3 hidden layers of 784, 128 and 64 units each with ReLU activation. **CNN Small** is a small convolutional neural network (Krizhevsky et al., 2012) with 1 convolution layer, 1 max pooling layer, 1 hidden layers of 128 units with ReLU activation. **CNN Large** is a large convolutional neural network (Krizhevsky et al., 2012) with 2 convolution layers, 2 max pooling layers, 2 hidden layers of 128 and 50 units each with ReLU activation.

1.2. Models for SST dataset

For the SST dataset, we use GloVe 300 dimensional word embeddings (Pennington et al., 2014) to convert words into vectors. A sequence length of 56 is selected and shorter sequence are zero-padded on the left. Longer sequences are truncated. We implement the following models:

LR is a Logistic Regression classifier (Collins et al., 2002). **DAN Small** is a small non-linear deep averaging network (Iyyer et al., 2015) that performs averaging of temporal features followed by a hidden layer of 64 units with 0.5 dropout and with ReLU activation. **DAN Large** is a small non-linear deep averaging network (Iyyer et al., 2015) that performs averaging of temporal features followed by 3 hidden layers of 128, 64 and 32 units each with 0.5 dropout and ReLU activation. **LSTM Small** is a small Long Short Term Memory network (Hochreiter and Schmidhuber, 1997) with 128 units followed by a hidden layer of 32 units with 0.5 dropout and ReLU activation. **LSTM Large** is a large 2 layer Stacked Bidirectional Long Short Term Memory network (Graves et al., 2013; Schuster and Paliwal,

[1997](#)) with 128 units followed by 2 hidden layers of 64 and 32 units each with 0.5 dropout and ReLU activation.

References

Michael Collins, Robert E. Schapire, and Yoram Singer. Logistic regression, adaboost and bregman distances. *Mach. Learn.*, 48(1-3):253–285, September 2002. ISSN 0885-6125. doi: 10.1023/A:1013912006537. URL https://doi.org/10.1023/A:1013912006537.

A. Graves, A. r. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6645–6649, May 2013. doi: 10.1109/ICASSP.2013.6638947.

Simon Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition, 1998. ISBN 0132733501.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9 (8):1735–1780, November 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL http://dx.doi.org/10.1162/neco.1997.9.8.1735.

Mohit Iyyer, Varun Manjunatha, Jordan L Boyd-Graber, and Hal Daumé III. Deep unordered composition rivals syntactic methods for text classification. In *ACL (1)*, pages 1681–1691, 2015.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012. URL http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *EMNLP*, volume 14, pages 1532–1543, 2014.

M. Schuster and K.K. Paliwal. Bidirectional recurrent neural networks. *Trans. Sig. Proc.*, 45(11):2673–2681, November 1997. ISSN 1053-587X. doi: 10.1109/78.650093. URL http://dx.doi.org/10.1109/78.650093.