Addendum - Robustness of sparse MLPs for supervised feature selection

Neil Kichler
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
n.j.kichler@student.utwente.nl

ABSTRACT

This addendum provides additional results for the robustness of sparse MLPs for supervised feature selection. In particular, feature selection using SET-MLP is performed on the Madelon dataset. We can observe that SET-MLP does not converge to a good feature selection resulting in an accuracy between 50-60% for support-vector, k-nearest neighbour and extra-trees classifier. It shows some limitations of this method.

1. ADDITIONAL RESULTS

This section presents the findings of the experiments for the Madelon dataset [1]. It is a binary classification task generated from artificial random data. Each sample contains 5 informative features and 15 linear combinations of the informative features. The noisy features act as distraction and should optimally get filtered out. The same hyperparameters are used as in the original paper for SET-MLP. Most importantly, the test size is $\frac{2}{3}$ for testing and the rest is for validation. The sparsity level of SET is $\epsilon = 13$.

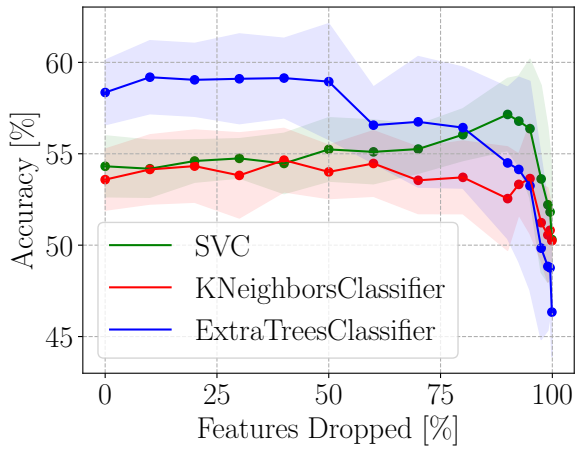

Figure 1. Classification accuracy of support-vector (green), k-nearest (red) and extra-trees (blue) classifier for different feature selections. The feature selection is performed using the proposed method. 12 runs, Madelon.

35th Twente Student Conference on IT July 2nd, 2021, Enschede, The Netherlands.

Looking at Figure 1, the general accuracy is significantly worse than the other datasets given the nature of Madelon. For SVC, we can see the mentioned peculiarities of Madelon in that it can further increase the accuracy of the method if more features are discarded. The other methods do not seem to show this behaviour, however.

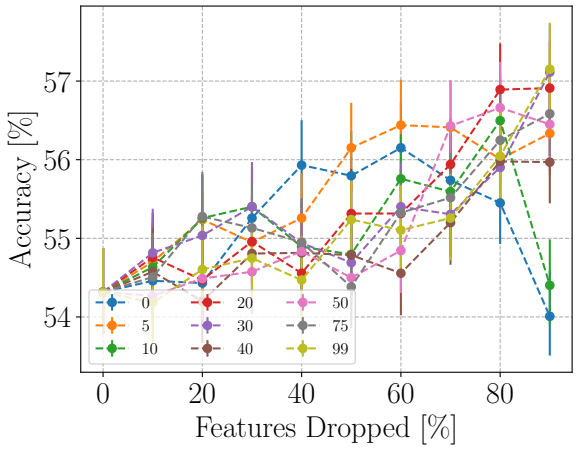

Figure 2. Support-Vector classifier accuracy for different feature selections. The legend shows the colours of the different training epochs of SET-MLP before the feature selection was performed. Based on 12 runs, Madelon.

In Figure 2, we can see the mentioned peculiarities of Madelon in that it can further increase the accuracy of the method if more features are discarded. However, the other classifiers do not show this behaviour (see below). Again, we can observe that the worst performers at 90% features dropped are initially worse for epoch 0 and 10 (interestingly not epoch 5). Also interesting is that between 40-60% features dropped the initial network actually beats the trained one. Overall, the accuracy is low with 54-57%.

In Figure 3 and 4, unlike SVC, we can observe a drop in accuracy with the percentage of features dropped. Here the difference between the feature selection of the initialized and trained SET-MLP becomes more apparent. The accuracy of the Extra-Trees classifiers is higher at its peak with 60%.

The variance of the k-nearest neighbour classifier is higher, but the general characteristics of epoch 0 performing worse can also be seen here.

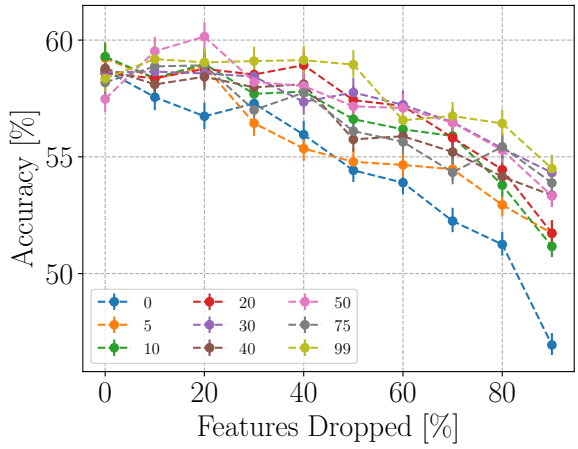

Figure 3. Extra-Trees classifier accuracy for different feature selections. The legend shows the colours of the different training epochs of SET-MLP before the feature selection was performed. Based on 12 runs, Madelon.

2. DISCUSSION

These results show the overall lack of convergence of SET for the specific Madelon dataset. The maximum accuracy SET-MLP itself achieves in classifying Madelon is 58.5% in our experiments and seems to be insufficient to provide meaningful feature selection. However, this does not mean that feature selection with SET-MLPs does not work at all, given the results of the original paper. It merely shows that when SET is too slow to converge for the Madelon dataset, resulting in poor feature selection performance.

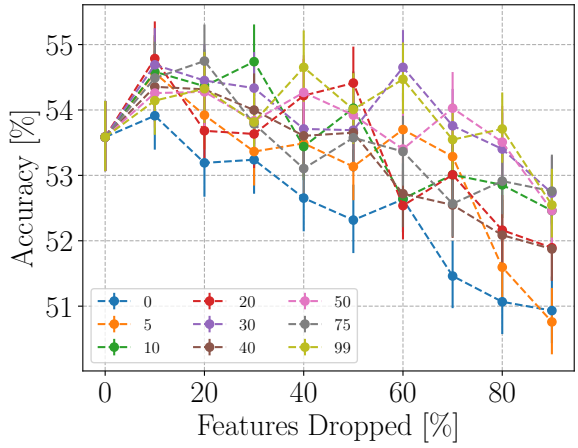

Figure 4. K-Nearest Neighbours classifier accuracy for different feature selections. The legend shows the colours of the different training epochs of SET-MLP before the feature selection was performed. Based on 12 runs, Madelon.

3. CONCLUSION

In light of these results, we can observe some limitations of SET-MLPs for feature selection. It seems that SET cannot converge and thus cannot discriminate the useful features from the redundant features.

References

[1] Isabelle Guyon. 2008. Madelon. UCI Machine Learning Repository. `http://archive.ics.uci.edu/ml/datasets/Madelon`