

Machine Learning Project Proposal

Praful Agrawal, Michael Matheny

February 11, 2015

A basic assumption in most machine learning is that the order of features in a feature vector is preserved in training as well as in the test instances. This is essential for obvious reasons of correspondence. However, it may happen that the feature values get permuted, a simple example of this scenario could be say if we have a feature vector $v = [a_1, a_2, a_3, a_4]$, but through some incident it has been reordered to be $v' = [a_4, a_1, a_2, a_3]$. Classifiers that are invariant to feature vector permutation could have critical medical applications such as epilepsy detection using EEG signals. In the standard procedure, electrical activity across the skull is recorded through many small metal disks. The resulting signals are processed using signal processing and then machine learning techniques are used to estimate the presence or impact of epilepsy. In different hospital settings (commonly observed in developing nations), if the disk numbers are not in correspondence with the desired input stream of a machine learning algorithm, the recommendations of such a system would be irrelevant. In such a critical scenario, recapture of data is also impractical, therefore, it would be helpful if machine learning algorithms become invariant to such permutations.

Developing a classifier that is invariant to all permutation will significantly increase the complexity of the classifier as well as the classification error. In this project, we aim to look at this problem of permuted feature vectors. It is a challenging problem for machine learning algorithms as it is a violation of very basic assumptions. The problem has not been addressed in the machine learning literature with much emphasis. One of the possible and interesting directions is proposed in [MCS10] where the idea is to update the classifier mapping to handle such permuted cases. Another direction could be to check if the given feature vector is in a valid order. This approach could also be helpful in analysis of outlier cases and assist the classifiers to operate on such cases with better performance.

References

- [MCS10] D. Roth M. Chang, D. Goldwasser and V. Srikumar. Discriminative learning over constrained latent representations. In *Proc. of the Annual Meeting of the North American Association of Computational Linguistics (NAACL)*, 6 2010.