Aspects of Online Learning

Edward Francis Harrington

6 December 2004

A thesis submitted for the degree of
Doctor of Philosophy at
The Australian National University

Research School of Information Sciences and Engineering

Declaration

Except where otherwise indicated, this thesis is my own original work. Some of the material in thesis has already been published elsewhere in collaboration with others, the details of which are:

1. Harrington E., Kivinen J. and Williamson R. C., Channel Equalization and the Bayes Point Machine, International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 4, pp. 493-496, 2003.

2. Harrington, E., Kivinen, J., Williamson, R. C., Herbrich, R., & Platt J. Online Bayes Point Machine. Proceedings of PAKDD-03, 7th Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 241-252, Heidelberg: Springer Verlag, 2003.

3. Harrington E., Expert mixture methods for adaptive channel equalization, Kaynak, O. et al. (Eds.): International Conference of Artificial Neural Networks and Neural Information Processing (ICANN/ICONIP), LNCS 2714, pp. 538-545, Springer Verlag, 2003.

4. Harrington E. F., Online Ranking/Collaborative Filtering using the Perceptron Algorithm, Proceedings of the Twentieth International Conference of Machine Learning (ICML-2003), pp. 250-257, Morgan Kaufmann.

Edward Harrington

6 December 2004

Computer Sciences Laboratory
Research School of Information Sciences and Engineering
The Australian National University

Acknowledgements

My warmest thanks must go to my wife Annie, who provided me with great support and love, especially whilst writing this thesis.

I have been fortunate to have had the supervision of Bob Williamson, the many “fire side” like chats with him have kept my on track to complete this thesis. His guidance in machine learning and signal processing have proven to be invaluable. I would like to thank a number of people in machine learning community for their help over last three or so years: Peter Bartlett, Ralf Herbrich, Jyrki Kivinen, Shahar Mendelson, John Platt, Gunnar Rätsch, and Alex Smola. I am very grateful to Jyrki Kivinen for his supervision and his many thoughtful comments concerning this thesis which have lead to its improvement. I am also grateful to Ralf Herbrich for his guidance and encouragement throughout the last few years. I would like to thank the examiners for their constructive suggestions, as they resulted in a much improved thesis.

I am very grateful to my employer, the Defence Science and Technology Organisation, for their financial support in allowing me to undertake this PhD. Several people at DSTO have provide me with assistance during my PhD: Ian Doherty, Geoff Latham, Nigel McGinty and Garry Newsam. I am especially appreciative of the encouragement and the support given to me by Ian Doherty, and the late Geoff Latham, when applying for study leave to undertake a PhD. I would also like to thank Garry Newsam for his support during the later stages of my PhD.

Thanks to the head of Computer Sciences Laboratory, John Lloyd, for funding some of my travel during 2003. To Michelle Moravec many thanks go to her for her help with all my administration needs whilst at the Computer Sciences Laboratory.

The friendship of my fellow PhD students has helped me during the last three, or so years: Evan Greensmith, Omri Guttiman, Kee Siong Ng and Cheng Soon Ong. I also acknowledge the proof reading efforts of Evan Greensmith.

Lastly, I would like to thank my parents and family for their love and support.

Abstract

Online learning algorithms have several key advantages compared to their batch learning algorithm counterparts: they are generally more memory efficient, and computationally more efficient; they are simpler to implement; and they are able to adapt to changes where the learning model is time varying. Online algorithms because of their simplicity are very appealing to practitioners. This thesis investigates several online learning algorithms and their application. The thesis has an underlying theme of the idea of combining several simple algorithms to give better performance. In this thesis we investigate: combining weights, combining hypothesis, and (sort of) hierarchical combining.

Firstly, we propose a new online variant of the *Bayes point machine* (BPM), called the *online Bayes point machine* (OBPM). We study the theoretical and empirical performance of the OBPM algorithm. We show that the empirical performance of the OBPM algorithm is comparable with other large margin classifier methods such as the *approximately large margin algorithm* (ALMA) and methods which maximise the margin explicitly, like the *support vector machine* (SVM). The OBPM algorithm when used with a parallel architecture offers potential computational savings compared to ALMA. We compare the test error performance of the OBPM algorithm with other online algorithms: the Perceptron, the voted-Perceptron, and Bagging. We demonstrate that the combination of the voted-Perceptron algorithm and the OBPM algorithm, called voted-OBPM algorithm has better test error performance than the voted-Perceptron and Bagging algorithms. We investigate the use of various online voting methods against the problem of ranking, and the problem of collaborative filtering of instances. We look at the application of online Bagging and OBPM algorithms to the telecommunications problem of channel equalization. We show that both online methods were successful at reducing the effect on the test error of label flipping and additive noise.

Secondly, we introduce a new mixture of experts algorithm, the *fixed-share hierarchy* (FSH) algorithm. The FSH algorithm is able to track the mixture of experts when the switching rate between the best experts may not be constant. We study the theoretical aspects of the FSH and the practical application of it to adaptive equalization. Using simulations we show that the FSH algorithm is able to track the best expert, or mixture of experts, in both the case where the switching rate is constant and the case where the switching rate is time varying.

Contents

List of Figures

List of Tables