DD2434 Projects

Jens Lagergren

December 12, 2016

Abstract

The task of the project is to reproduce the results presented in a published scientific article, describe the article orally and in written form to your peer students, and argue for and against the method presented in the article. From this you will learn how to read scientific articles, how to implement and use a particular method, how to argue for and against a method, and how to adapt the presentation of a method to different target groups (i.e., adapt the presentation of the method in the article - targeted to active researchers in Machine Learning - so that it is understandable to first year Master students in Machine Learning). The below 4 papers represent a range of different topics in Machine Learning, and have been selected by Jens, who will be the supervisor of these projects.

Some of the papers are more theoretical and while others are of a more practical nature. The requirements will change accordingly, so if you pick a more practical paper you will need to perform more experiments while a more theoretical paper requires you to show a more thorough analysis of the paper. Detailed instructions about the project can be found on the course home page, Project in the menu to the left.

1 Graphical Models

M Meila and M Jordan. "Learning with mixtures of trees". In: *The Journal of Machine Learning Research* (2001)

In this paper Meila and Jordan extend a classical result by Chow and Liu. Chow and Liu showed in 1968 that a tree DGM can be learned, or rather that an ML estimate of the tree can be obtained. In this paper, Meila and Jordan introduce mixtures of trees (i.e., tree DGMs) and provide an EM algorithm for the ML estimation problem.

N Friedman. "The Bayesian structural EM algorithm". In: *Proc. UAI* (1998)

This is a classical machine learning paper where Nir Friedman shows that the EM algorithm can also be applied in order to maximize with respect to discrete parameters. The focus here is on finding a DGM with observable variables given data (for all those variables).

2 Variational Inference

Arto Klami *et al.* "Group factor analysis". In: *IEEE transactions on neural networks and learning systems* 26.9 (2015), pp. 2136–2147

This paper is an extension to Factor Analysis learning representations between groups of variables. This can for example be seen as learning from multimodal data, e.g., a set of variables represents text and another set represents images. In this paper, the learning is based on a variational approximation.

3 Clustering using a Dirichlet process

Andrew Roth *et al.* "PyClone: statistical inference of clonal population structure in cancer". In: *Nature Methods* 11.4 (2014), pp. 396–398

Shah (previous student of Kevin Murphy) and coworkers applies Dirichlet process based clustering to an important problem in analysis of cancer data. The model is complex and you better simplify it or the analysis.

References

- N Friedman. “The Bayesian structural EM algorithm”. In: *Proc. UAI* (1998).
- M Meila and M Jordan. “Learning with mixtures of trees”. In: *The Journal of Machine Learning Research* (2001).
- Arto Klami *et al.* “Group factor analysis”. In: *IEEE transactions on neural networks and learning systems* 26.9 (2015), pp. 2136–2147.
- Andrew Roth *et al.* “PyClone: statistical inference of clonal population structure in cancer”. In: *Nature Methods* 11.4 (2014), pp. 396–398.