DD2434 Projects

Carl Henrik Ek

December 5, 2014

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 7 papers represent a range of different topics in Machine Learning, and have been selected by Carl Henrik, 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 Sparse Modeling

Michael E Tipping. "The Relevance Vector Machine". In: *Advances in Neural Information Processing Systems* 12 (2000). URL: http://scholar.google.com/scholar?q=related:asJ-P2dN9hEJ:scholar.google.com/&hl=en&num=20&as_sdt=0,5

This paper, even though a bit out-dated, is an excellent introduction to sparse learning. Tipping is an excellent story-teller and there is lots of available material where he motivates Bayesian modeling in general and sparse Bayesian modeling in specific. The paper is a good mix between theory and practical applications but you should be fairly happy about using EM for this paper to go down smoothly.

E Snelson and Zoubin Ghahramani. "Sparse Gaussian processes using pseudo-inputs". In: *Advances in Neural Information Processing Systems* 18 (2006), p. 1257

Gaussian processes are quite expensive to learn due to how the kernel matrix of the training data appears inside an inverse in the marginal likelihood. In order to make learning more efficient there have been several suggestions of how to make the sparse approximations to GPs and this paper outlines such an approach. There are some quite nasty gradient computations in here which are not fun to do but otherwise the implementation is rather straight-forward. This paper is more of a theoretical work and you need to fully understand Gaussian processes to digest the work.

2 Representation Learning

Michael E Tipping and Christopher M Bishop. "Probabilistic principal component analysis". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.3 (1999), pp. 611–622. URL: http://onlinelibrary.wiley.com/doi/10.1111/1467-9868.00196/abstract

This paper is the original PPCA paper. In the assignment we looked briefly at this algorithm but now we want to look at it from a fully Bayesian perspective. The learning in this paper is done using the EM algorithm so if you think that is fun this might be the paper for you. The important part of the paper is the theoretical analysis so you do not have to run extensive experiments but rather should analyse what the model actually does and what it allows you to do.

B Schölkopf and A J Smola. "Kernel principal component analysis". In: *Artificial Neural Networks—ICANN'97* (1997). URL: http://www1.cs.columbia.edu/~cleslie/cs4761/papers/scholkopf_kernel.pdf

This is not a probabilistic paper which makes the implementation very easy but the analysis of the algorithm quite hard. If you choose this paper you should pick several different kernels and try a few different data-sets. You can think of the paper as a means of visualising the image of the kernel induced feature space, therefore you should try and analyse what happens with different kernels and different parameters of kernels for a specific set of data.

3 Gaussian Processes

Neil D Lawrence. "Probabilistic non-linear principal component analysis with Gaussian process latent variable models". In: *The Journal of Machine Learning Research* 6 (2005), pp. 1783–1816. URL: http://dl.acm.org/citation.cfm?id=1194904

This paper describes a dual formulation of PPCA which uses Gaussian process priors to marginalise out the generating mapping. The paper should be very straight-forward to implement except for some rather nasty gradients. The important thing in this paper is the analysis of the algorithm and explain why this actually works at all. You will also need to understand the PPCA paper and compare these different algorithms together.

4 Kernels

Kilian Q Weinberger *et al.* "Learning a kernel matrix for nonlinear dimensionality reduction". In: *International Conference on Machine Learning*. ACM, July 2004, pp. 106–113. ISBN: 1-58113-838-5. DOI: 10.1145/1015330.1015345. URL: http://portal.acm.org/citation.cfm?id=1015330.1015345

This paper is a very neat small and nice kernel learning paper. The formulation is very intuitive and the learning is quite simple as the problem is formulated as convex program. You should not implement your own solver but rather use existing packages. If you choose this paper you need to re-produce the results in the paper and provide an analysis of what the algorithm actually does. The reasoning is all geometric and can be done very intuitively.

H Lodhi *et al.* "Text classification using string kernels". In: *The Journal of Machine Learning Research* 2 (2002), pp. 419–444

This paper presents a kernel for sequences, specifically discrete sequences such as strings. This allows you to put things that do not live in vector spaces into such which is very neat and an important benefit of kernels. The paper is very nice and the derivation of the kernel excellent. Implementation of this work should be straight-forward. You will also need to understand Kernel PCA to be able to visualise the image of the feature space from data.

References

[] Michael E Tipping. "The Relevance Vector Machine". In: *Advances in Neural Information Processing Systems* 12 (2000). URL: http://scholar.google.com/scholar?q=related:asJ-P2dN9hEJ:scholar.google.com/&hl=en&num=20&as_sdt=0,5.

[] E Snelson and Zoubin Ghahramani. "Sparse Gaussian processes using pseudo-inputs". In: *Advances in Neural Information Processing Systems* 18 (2006), p. 1257.

[] Michael E Tipping and Christopher M Bishop. "Probabilistic principal component analysis". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.3 (1999), pp. 611–622. URL: http://onlinelibrary.wiley.com/doi/10.1111/1467-9868.00196/abstract.

[] B Schölkopf and A J Smola. "Kernel principal component analysis". In: *Artificial Neural Networks—ICANN'97* (1997). URL: http://www1.cs.columbia.edu/~cleslie/cs4761/papers/scholkopf_kernel.pdf.

[] Neil D Lawrence. "Probabilistic non-linear principal component analysis with Gaussian process latent variable models". In: *The Journal of Machine Learning Research* 6 (2005), pp. 1783–1816. URL: http://dl.acm.org/citation.cfm?id=1194904.

[] Kilian Q Weinberger *et al.* "Learning a kernel matrix for nonlinear dimensionality reduction". In: *International Conference on Machine Learning*. ACM, July 2004, pp. 106–113. ISBN: 1-58113-838-5. DOI: 10.1145/1015330.1015345. URL: http://portal.acm.org/citation.cfm?id=1015330.1015345.

[] H Lodhi *et al.* "Text classification using string kernels". In: *The Journal of Machine Learning Research* 2 (2002), pp. 419–444.