# The list of four papers proposed by Pawel for a group project in DD2434 (mladv19)

## 1. Sparse Modelling

M.E. Tipping (2001) Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research* 1.

M.E. Tipping (2000) The Relevance Vector Machine. *Advances in Neural Information Processing Systems* 12.

These papers, even though a bit outdated, are an excellent introduction to sparse learning. Tipping is an excellent story-teller, who motivates Bayesian modelling in general and sparse Bayesian modelling, in particular. The paper (Tipping, 2000) is a good mix between theory and practical applications but you should be satisfied with using EM for this paper.

Your project could focus on a subset [1] of the following tasks

- implement the proposed methods, validate Tipping's results on a toy problem
- deploy RVM on selected regression and classification problems, and compare the results and interpretation to SVM
- compare, interpret probabilistic outputs obtained with RVMs and SVMs (look up research on probabilistic output of SVMs)
- investigate scaling properties of RVMs (compare with SVMs)
- critically examine/discuss Tipping's arguments in favour of RVMs
- review the literature on RVMs – mention what other algorithms have been proposed based on RVMs and identify a few seminal examples where RVMs have been successfully applied.

[1] As mentioned in the instructions on the webpage, a constellation of tasks and the depth of their implementation/execution constitute the main criteria for project assessment. This applies to all papers.

## 2. Kernels

O. Chapelle, J. Weston, and B. Schölkopf (2003) Cluster kernels for semi-supervised learning. *Advances in neural information processing systems,* pp. 601-608.

In short, this paper demonstrates how the notion of similarity between data samples mediated by the kernel concept can be directly exploited for clustering. It also brings up a important class of semi-supervised learning problems, which currently receive a lot of attention. The paper illustrates the design of clusters implementing the cluster assumption and reports the results of their comparative evaluation with other methods, e.g. support vector machines (SVMs), on different benchmark tests.

Here are suggested tasks that could constitute your project

- re-implement a subset of the proposed kernels, reproduce the selected benchmark results and apply to new problems; test scaling properties of the kernel-based approach in comparison with other methods included in the paper
- make a comparative evaluation of different kernels on a subset of benchmark tasks (and potentially new problems of your choice) including other machine learning methods than SVMs, for example relevance vector machines (RVMs)
- study the semi-supervised nature of your learning problem with respect to, for example, the number of available labels; for new benchmark problems of your choice, how does the performance of the proposed kernel based approach depends on kernel parameters for varying proportion of labelled samples? How is it affected by the choice of labelled vs unlabelled data points?
- propose your own kernels and conduct a systematic comparative evaluation on a selected subset of benchmark tests.

## 3. Gaussian Processes and Representation Learning

Neil D. Lawrence (2005) Probabilistic non-linear principal component analysis with Gaussian process latent variable models". *The Journal of Machine Learning Research* 6, pp. 1783-1816.

Neil D. Lawrence (2004) Gaussian process models for visualisation of high dimensional data. In S. Thrun, L. Saul, and B. Scholkopf, editors, *Advances in Neural Information Processing Systems*, vol. 16, pp. 329–336, Cambridge, MA, MIT Press.

These papers introduce the concept of Gaussian process latent variable model (GP-LVM) that can be considered as a multiple-output GP regression model where only the output data are given. In this respect the model implements representation learning with the unobserved inputs treated as latent variables, which are optimised instead of being integrated out. This perspective renders the model as a nonlinear extension of the linear probabilistic principal component analysis (PPCA). In the other paper GPLVM is then evaluated as an approach to the visualisation of high dimensional data.

In the project you could examine in depth and build on the following tasks

- reimplement the proposed method by following model development in the paper including the MAP estimate and its maximization for hyperparameter identification
- demonstrate the use of the model for visualization of new high-dimensional datasets
- make a comparative evaluation of GPLVM with a kernel PCA algorithm on selected high-dimensional datasets
- extend the proposed GPLVM algorithm by devising a variational Bayesian approach to the marginalization of latent (input) variables, which would optimize the lower bound on the marginal likelihood wrt hyperparameters
- another option for extension could be model sparsification
- elaborate further on the discussion points in Lawrence (2004).

You are requested to develop your own implementations rather than rely on any existing libraries, particularly the code referred to in the papers.