The list of DD2434 papers for projects under Pawel Herman's supervision (adapted from the 2015 version proposed originally by C.H. Ek)

December 2016

1. Sparse Modelling

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

This paper, even though a bit outdated, is an excellent introduction to sparse learning.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

E. Snelson and Z. Ghahramani (2006) Sparse Gaussian processes using pseudo-inputs. *Advances in Neural Information Processing Systems* 18.

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 at least the implementation is rather straightforward. This paper is more of a theoretical work and you need to fully understand Gaussian processes to digest the work.

2. Representation Learning

M.E. Tipping and C.M. Bishop (1999) Probabilistic principal component analysis. *Journal of the Royal Statistical Society*: Series B (Statistical Methodology) 61.3, pp. 611–622

This paper is the original PPCA paper. We have looked briefly at this algorithm but now we want to examine it from a fully Bayesian perspective. The learning in this paper is done using the EM algorithm. The important part is the theoretical analysis so you do not have to run extensive experiments but should rather focus on what the model actually does and what it allows you to do.

3. Kernels

H. Lodhi, C. Saunders, J. Shawe-Taylor, N. Cristianini, and C. Watkins (2002) Text classification using string kernels. *J. Mach. Learn. Res.* 2, 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.