Classification of fMRI data using latent Gaussian models

Maurizio Filippone    Mark Girolami

Department of Statistical Science
University College London

January 24th, 2011
The problem

- Infer subject’s cognitive state from fMRI data
- Discriminate between cognitive states as well as constructing multivariate brain maps (which brain regions carry discriminative information)
- Linear SVMs and Bayesian logistic regression have been applied with success (Mourão-Miranda 2005 et al., Marquand et al. 2010)
The problem

- Infer subject’s cognitive state from fMRI data
- Discriminate between cognitive states as well as constructing multivariate brain maps (which brain regions carry discriminative information)
- Linear SVMs and Bayesian logistic regression have been applied with success (Mourão-Miranda 2005 et al., Marquand et al. 2010)
- Fully Bayesian non-linear discriminative method
- Classifiers based on Gaussian Processes are one instance of latent Gaussian models
Latent Gaussian Models - (LGM)

<table>
<thead>
<tr>
<th>$p(\theta)$</th>
<th>prior $\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K = LL^T$</td>
<td>covariance matrix</td>
</tr>
<tr>
<td>$\nu \sim \mathcal{N}(0, I)$</td>
<td>whitened latent</td>
</tr>
<tr>
<td>$f = L\nu$</td>
<td>transformation</td>
</tr>
<tr>
<td>$\nabla p(f</td>
<td>\theta) = \mathcal{N}(f</td>
</tr>
<tr>
<td>$p(y</td>
<td>f) = \mathcal{E}(y</td>
</tr>
</tbody>
</table>

Squared exponential covariance function

$$k(x_i, x_j|\theta) = \alpha \exp \left[-\frac{1}{2}(x_i - x_j)^T A(x_i - x_j)\right]$$
LGM - Logistic regression example
LGM - Logistic regression example
• Log-Gaussian Cox model (Møller et al. 1998)
• Gaussian copula process volatility model (Wilson and Ghahramani 2010)
• Gaussian processes for ordinal regression (Chu and Ghahramani 2005)
Bayesian inference in LGM

Why Bayesian?

A fully Bayesian approach provides a way of:

- including prior information
- inferring model parameters
- obtaining predictive distributions (balance cost of decisions)
- approaching online learning
- doing model selection

Bayesian inference for these models is intractable
Markov Chain Monte Carlo (MCMC) methods provide a way to sample from the posterior distribution of the model parameters, but:

- computation of the likelihood is in $O(n^3)$ (same complexity for approximate methods)
- how to devise an efficient sampling mechanism? (e.g., what sampler, variable blocking, parametrization)
- conditional distributions $p(f|\theta, y)$ and $p(\theta|f, y)$ are such that Gibbs sampler updates require a Metropolis acceptance step
The structure of the model poses a serious challenge to MCMC methods for efficiently sampling from posterior distributions.
Centered vs non-centered parametrizations (Papaspiliopoulos et al. 2007)
Data

- Experiments reported here are with a single subject listening passively to vocal and non-vocal stimuli
- Preprocessing: time correction, spatial smoothing, masking, normalization, and voxel reduction ($t$-test)
- We have 200 samples with 4,436 covariates (number of voxels remaining after the $t$-test)
- classes: 1 vocal and 0 non-vocal stimuli
Results - Experimental setting

- classifier based on GP (GPC) (same cost for the two classes)
  - Gibbs sampler:
  - \( f|\theta, y \) using manifold methods
  - \( \theta|f, y \) using non-centered parametrization (i.e., \( \theta|\nu, y \))
- Support Vector Machines (SVM)
  - tested with both linear and radial basis function kernel
  - parameters (\( C \) and kernel bandwidth) were optimized using 10-fold cross validation
- GPC and non-linear SVMs use isotropic covariance/kernel functions
Results - Classification accuracy

Classification result using 4-fold validation

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy (std err)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM (lin)</td>
<td>75.5% (5.9%)</td>
</tr>
<tr>
<td>SVM (rbf)</td>
<td>76% (1.4%)</td>
</tr>
<tr>
<td>GPC</td>
<td>78.5% (3.8%)</td>
</tr>
</tbody>
</table>

- we can use the predictive distribution for finer decision rules
- by doing so we achieve 92.8% accuracy on 90 samples
Conclusions and ongoing work

• We are devising efficient sampling methods for full Bayesian inference in latent Gaussian models
• In the application to fMRI data, performance of the GP based classifier comparable to SVMs
• Benefits of a fully Bayesian treatment in the descriptive power of the model
• Include a posterior inference of covariates weights in the sampling mechanism
• Design of covariance/kernels for fMRI data
This research is funded by the Engineering and Physical Sciences Research Council (EPSRC), grant number EP/E052029/1.