Scope of this work

In clinical neuroimaging applications where subjects belong to one of multiple classes of disease states and multiple imaging sources are available, the aim is to achieve accurate classification while assessing the importance of the sources in the classification task.

Data

- Structural magnetic resonance imaging (MRI) data
  - T1-weighted structural imaging
  - Preprocessing using the SPMfMRI software package ([www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm))

Parkinsonian data

- Segmentation parcellating anatomically into six target regions of interest (brainstem, cerebellum, caudate, middle occipital gyrus, putamen, and one for all other brain regions)
- 62 subjects (healthy controls + patients with one of three akinetic-rigid neurological disorders)
  - 14 subjects healthy controls
  - 18 subjects multiple system atrophy (MSA)
  - 16 subjects progressive supranuclear palsy (PSP)
  - 14 subjects idiopathic Parkinson’s disease (IPD)

ADHD and ASD data

- 77 adolescent subjects (aged 10-18)
  - 29 subjects healthy controls
  - 29 subjects with either attention deficit/hyperactivity disorder (ADHD)
  - 19 subjects with autism spectrum disorder (ASD)

Methods

Classification approach

- Multiple class
- Kernel-based - pairwise similarity between subjects
- Multiple kernel - multiple input data
- Probabilistic - we need sound quantification of uncertainty

Multi-Class Multiple Kernel classifier using Gaussian Processes

Classification Model

Data

- \((x_1, y_1), \ldots, (x_n, y_n)\) set of n input/class pairs
- \(y_i = (y_{i1}, \ldots, y_{ik})\) with \(y_{ij}\) belonging to class \(j\)

Model

- Class labels conditionally independent given a set of class specific latent variables
- \(\mathbf{y} = (y_1, \ldots, y_k)^T\) with \(y_i^k = 1\) if \(x_i\) belongs to \(k\)th class
- \(\pi(y) = \prod_{i=1}^{n} \pi(y_i|x_i)\)

Inference and predictions

- Predictions of label \(y_i\) for new input \(x_i\)
- Approximate integral using Markov chain Monte Carlo
- \(\pi(y|x)\) is very inefficient
- Approximate the predictive distribution by
  \[p(x|y) = \frac{1}{N} \sum_{i=1}^{N} p(X_i|Y_i)\]

Asymptotically exact

Challenges

- Not possible to draw samples from \(p(\theta|y)\)
- Need to resort to Gibbs sampling types of schemes
- \(p(\theta|y)\) makes the MCMC approach very inefficient (sampling \(p(\theta|y)\) is relatively easy instead)

Proposal

- Sampling \(y\) using the Metropolis-Hastings algorithm
  - Initialize the algorithm randomly from \(\theta\)
  - Propose a new set of parameters \(\theta^{\prime}\) from \(p(\theta^{\prime}|y)\)
  - Accept proposal with probability

\[
\min \left\{ \frac{p(\theta^{\prime}|y)}{p(\theta|y)} \right\} \]

This will yield samples from the correct \(p(y|x)\)

We propose to estimate \(p(\theta|y)\) using importance sampling

\[
p(y|x) \approx p(y|x) = \frac{1}{N_{\text{import}}} \sum_{i=1}^{N_{\text{import}}} \frac{p(x_i|y)}{q(x_i|y)}
\]

where \(N_{\text{import}}\) samples are drawn from a Gaussian distribution \(q(\theta|y)\) approximating \(p(\theta|y)\)

- Variance of the estimator affects MCMC efficiency

Drawing from \(q(\theta|y)\)

Implementation requiring the storage of \(n \times n\) matrices only

Sequentially draw latent variables pertaining to each class

- Define precision of \(\theta^{\prime}\) as \(\Lambda + K^{-1} + \text{diag}(\pi) - \text{Hessian}\)
- Inverse covariance of \(\theta^{\prime}\)

\[
\Lambda_{\text{new}} = \Lambda - \Lambda_{\text{old}} \Lambda_{\text{old}}^{\dagger} \Lambda_{\text{old}}^{\dagger}
\]

- Can be inverted storing at most \(n \times n\) matrices
- Mean of \(\theta^{\prime}\)

Conclusions and ongoing work

- We proposed the use of Gaussian Process classification for multiple-class/multiple-kernel (MC-MKL) learning for neuroimaging data where:
  - Quantification of uncertainty in predictions is of primary interest
  - Assessment of importance of different sources of information is key to design future experiments
- We demonstrated that exact quantification of uncertainty in parameter estimates leads to better predictions compared to approximate methods
- We proposed a practical way to carry out exact quantification of uncertainty based on advanced Markov chain Monte Carlo methods
- A large variance for \(p(y|x)\) can severely reduce the efficiency of the proposed method
- We are studying alternatives to importance sampling and the Laplace Approximation to reduce the variance of \(p(y|x)\)