Pseudo-Marginal Bayesian Inference for Gaussian Processes
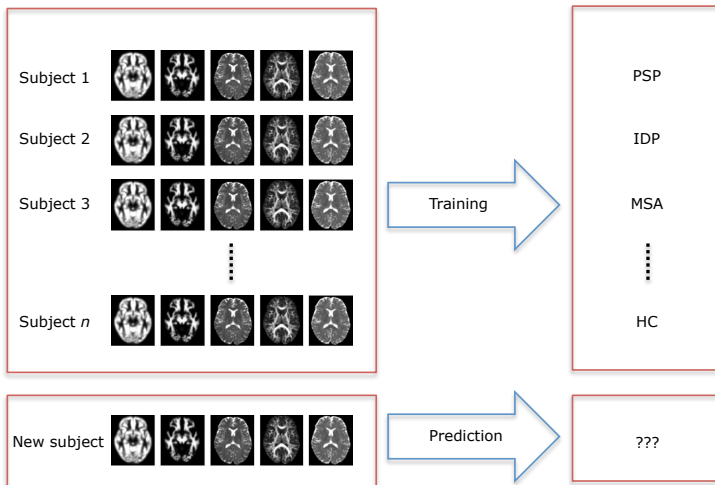
Maurizio Filippone

School of Computing Science
University of Glasgow
maurizio.filippone@glasgow.ac.uk

May 20th, 2014

Motivating Application

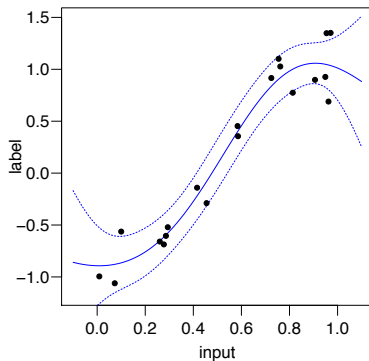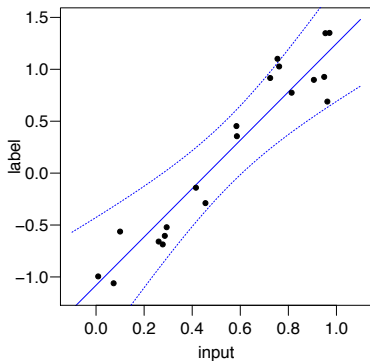HC - Healthy control

MSA - Multiple system atrophy

PSP - Progressive Supranuclear Palsy

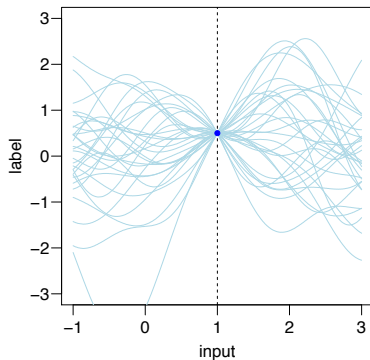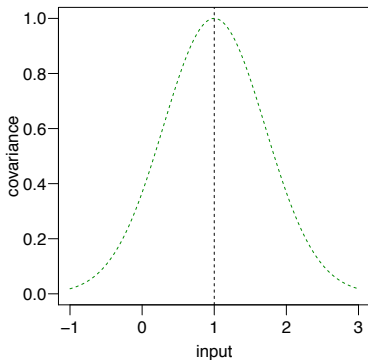IDP - Idiopathic Parkinson's disease

- Mapping input to labels

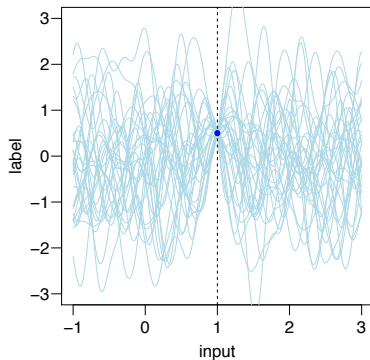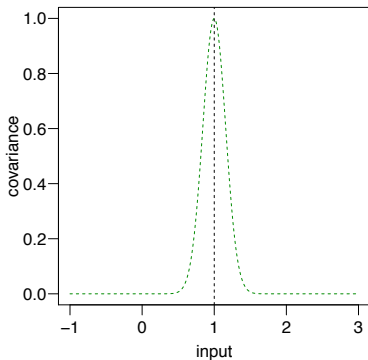$$\text{label} = \text{function}(\text{input}, \text{par}, \text{noise})$$

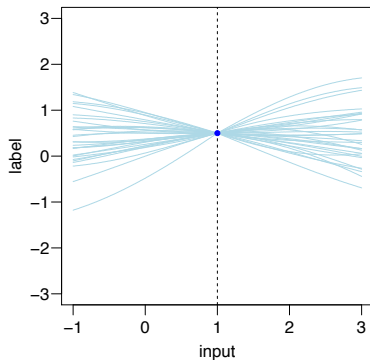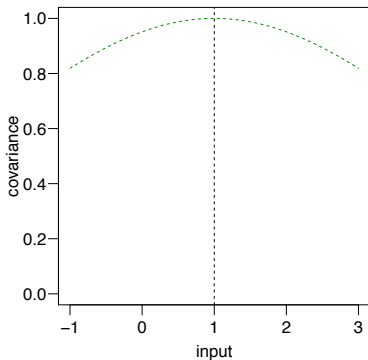- We may have no clue about function - we need assumptions

Gaussian Processes

- Gaussians with distance dependent covariance

Gaussian Processes

- Gaussians with distance dependent covariance

Gaussian Processes
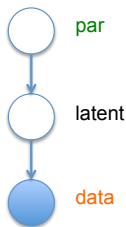
- Gaussians with distance dependent covariance
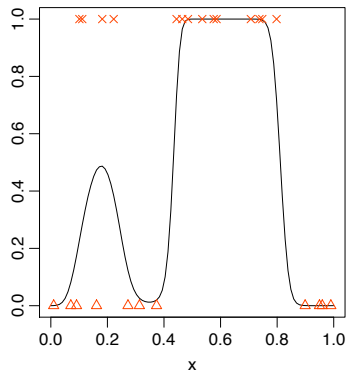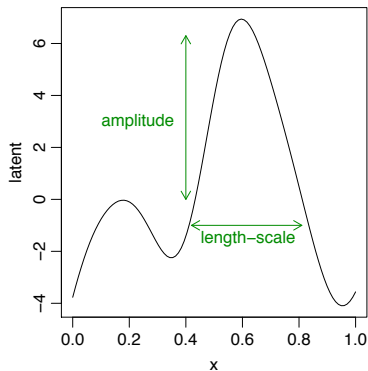
Gaussian Process Models

- Class of hierarchical models

$$p(\text{data}|\text{latent}) \qquad p(\text{latent}|\text{par}) \qquad p(\text{par})$$

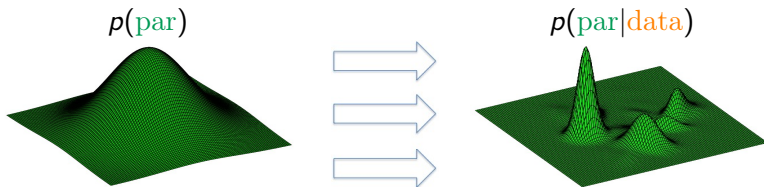- $p(\text{latent}|\text{par}) = \text{Gaussian Process}$

Gaussian Process Models - Classification example
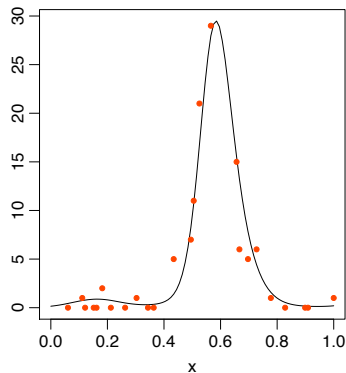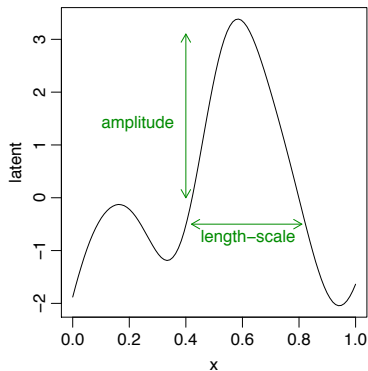
- Inference using Bayes theorem:

$$p(\mathrm{par}|\mathrm{data}) = \frac{p(\mathrm{data}|\mathrm{par})p(\mathrm{par})}{\int p(\mathrm{data}|\mathrm{par})p(\mathrm{par})d\mathrm{par}}$$

$p(\mathrm{par})$

$p(\mathrm{par}|\mathrm{data})$
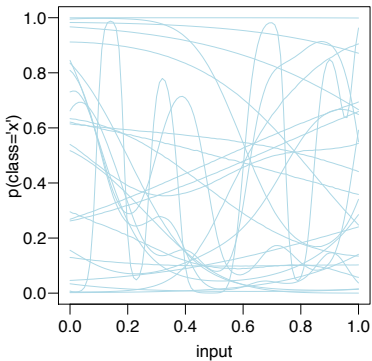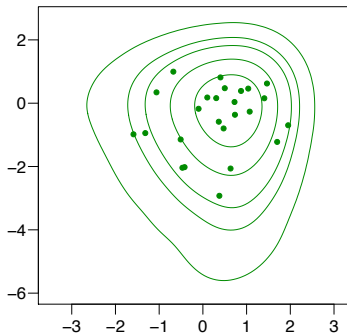
Bayesian Inference - Data
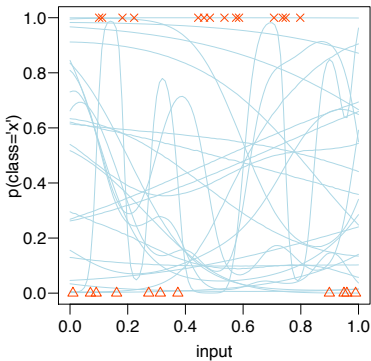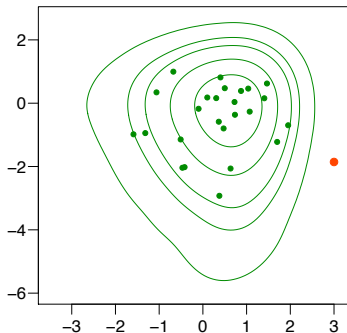
Bayesian Inference and Predictions
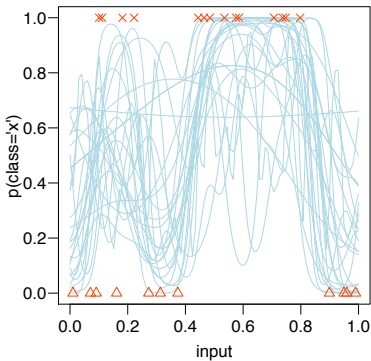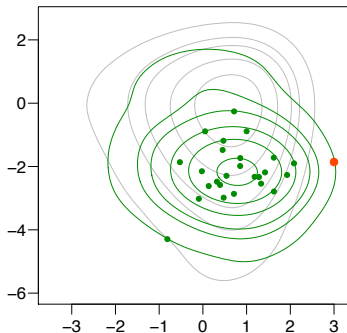
Bayesian Inference and Predictions
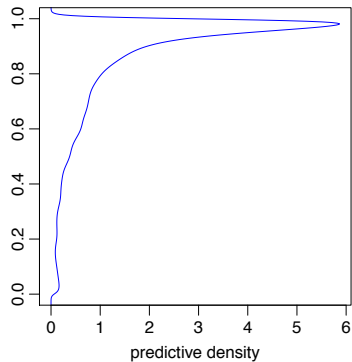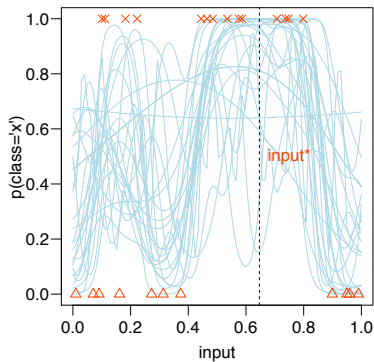
- Predictions for new data

$$p(\text{label}_* | \text{label}) = \int p(\text{label}_* | \text{par}) p(\text{par} | \text{label}) \, d\text{par}$$

- Monte Carlo integration:

$$\int p(\text{label}_* | \text{par}) p(\text{par} | \text{label}) \, d\text{par} \simeq \frac{1}{N} \sum_{i=1}^{N} p(\text{label}_* | \text{par}^{(i)})$$

with $\text{par}^{(i)}$ drawn from $p(\text{par} | \text{data})$

- Draw samples according to the posterior density
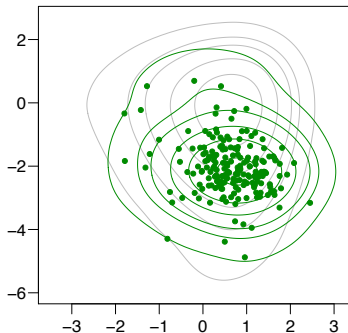
- Bayesian inference

$$p(\mathrm{par}|\mathrm{data}) = \frac{p(\mathrm{data}|\mathrm{par})p(\mathrm{par})}{\displaystyle\int p(\mathrm{data}|\mathrm{par})p(\mathrm{par})d\mathrm{par}}$$

- Random walk sampler - accept a proposal with probability

$$\min\left(1, \frac{p(\mathrm{par}'|\mathrm{data})}{p(\mathrm{par}|\mathrm{data})}\right)$$

- Explore the parameter space according to the density
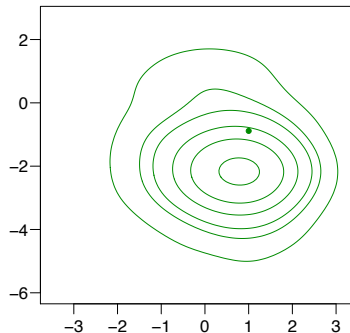
- Explore the parameter space according to the density

- Explore the parameter space according to the density

- Explore the parameter space according to the density
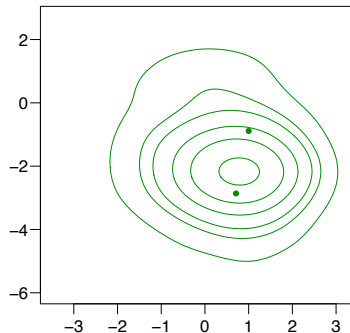
- Explore the parameter space according to the density
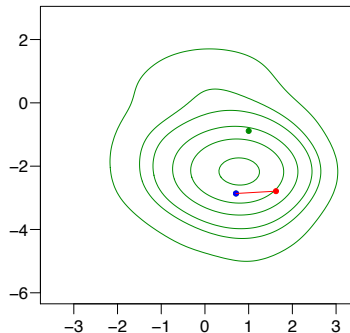
- Explore the parameter space according to the density
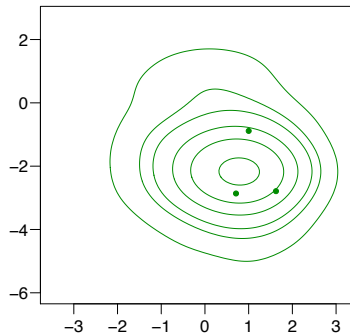
- Explore the parameter space according to the density

- Explore the parameter space according to the density

- Explore the parameter space according to the density

- Explore the parameter space according to the density
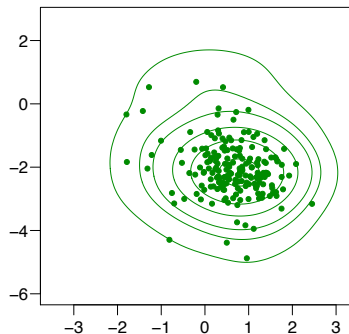
Marginal likelihood

$$p(\text{data}|\text{par}) = \int p(\text{data}|\text{latent})p(\text{latent}|\text{par})d\text{latent}$$

is unavailable analytically. Options:

- Approximate $p(\text{data}|\text{par})$ within MCMC
- Sample from $p(\text{par}, \text{latent}|\text{data})$
- Pseudo-Marginal MCMC

Gaussian approximation to $p(\text{latent}|\text{data}, \text{par})$

- Laplace Approximation
- Expectation Propagation
- Variational Bayes
- . . .

- No exact Gibbs steps - need to employ Metropolis within Gibbs steps - waste of computations when rejecting
- Updates of par cost $O(n^3)$
- par can be large dimensional (e.g., Automatic Relevance Determination (ARD) covariance function)
- There are n latent variables (as many as the number of observations)

Challenges in MCMC for GPMs - Structure

Obvious iterative scheme (aka Sufficient Augmentation (SA) scheme). Alternate between:

- Drawing from $p(\text{latent}|\text{par}, \text{data})$
- Drawing from $p(\text{par}|\text{latent})$ (**bad idea** - see figure)

Ancillary Augmentation (AA) scheme - reparametrization:

$$K = LL^{\mathrm{T}} \qquad \mathrm{ancillary} = L^{-1}\,\mathrm{latent}$$

- Replace sampling of par with $p(\mathrm{par}|\mathrm{ancillary}, \mathrm{data})$
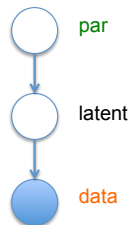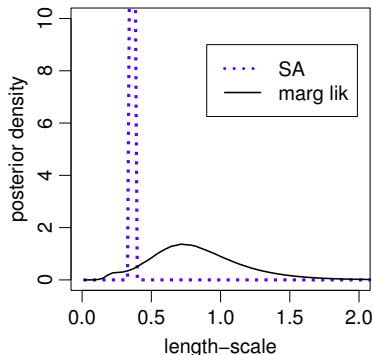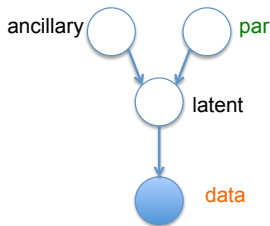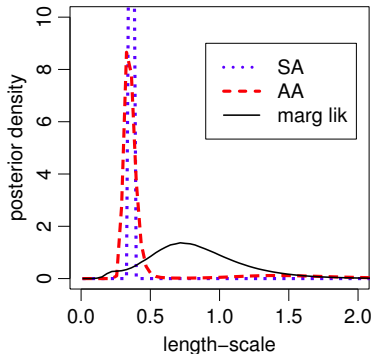
Mitigating coupling effect through reparameterization

Surrogate data model (SURR):

- Introduce set of auxiliary variables informed by the posterior over latent
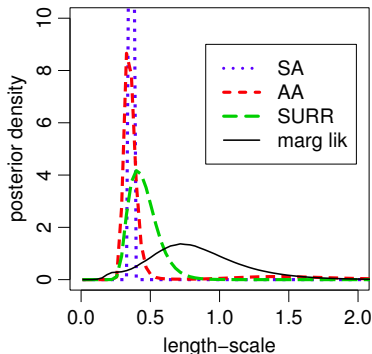
$$\mathrm{surrogate} = f(\mathrm{latent}, \mathrm{par})$$

- Replace posterior by unbiased estimate

$$\min \left(1, \frac{\tilde{p}(\text{par}'|\text{data})}{\tilde{p}(\text{par}|\text{data})} \right)$$

- Achieved by using an unbiased estimate of $\tilde{p}(\text{data}|\text{par})$

Importance Sampling estimator

- Approximate posterior over latent variables using $q(\mathrm{latent})$
- Then

$$\tilde{p}(\mathrm{data}|\mathrm{par}) = \frac{1}{N} \sum_{i=1}^{N} \frac{p(\mathrm{data}|\mathrm{latent}^{(i)})p(\mathrm{latent}^{(i)}|\mathrm{par})}{q(\mathrm{latent}^{(i)})}$$

Mixing
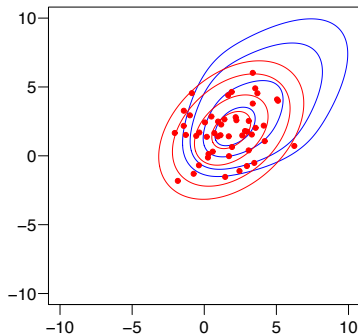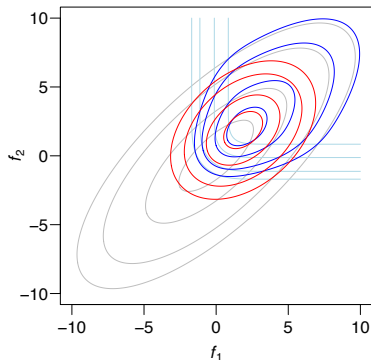
Abalone data set (two classes) $n = 2835$ - inference of length-scale

- Annealing from an approximation

- Annealing from an approximation
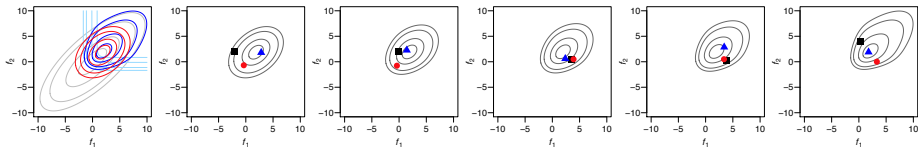
- Annealing from the prior
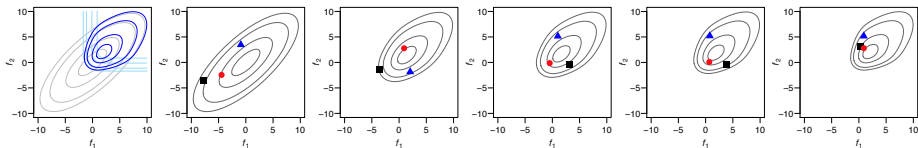
Analysis of the variance of the AIS and IS estimators

- r is the variance of the $\log_{10}$ marginal likelihood
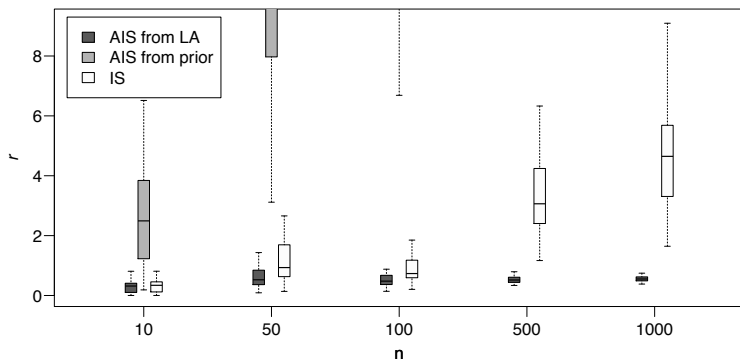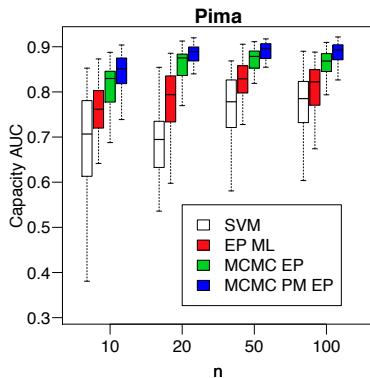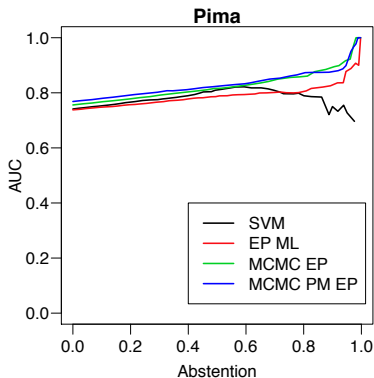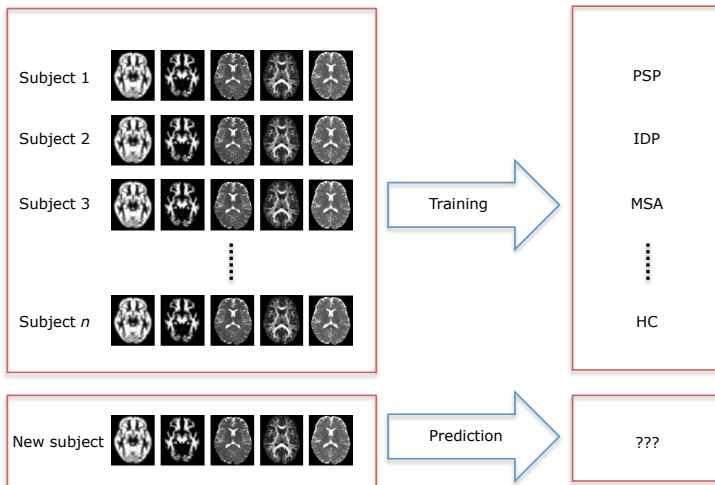
Motivating Application

Subject 1

Subject 2

Subject 3

⋮

Subject n

Training

PSP

IDP

MSA

⋮

HC

New subject

Prediction

???

HC - Healthy control

MSA - Multiple system atrophy

PSP - Progressive Supranuclear Palsy

IDP - Idiopathic Parkinson's disease

- Multiclass classification based on GPs

$$p(\text{disease} = c|\text{sources}) = \text{unknown function}$$

- unknown function modeled using GPs

- Covariance based on source-dependent covariances S_k

$$\sum_{k=1}^{K} w_{ck} S_k(subject_i, subject_j)$$

Method	Accuracy
GP classifier	0.598
SimpleMKL	0.418
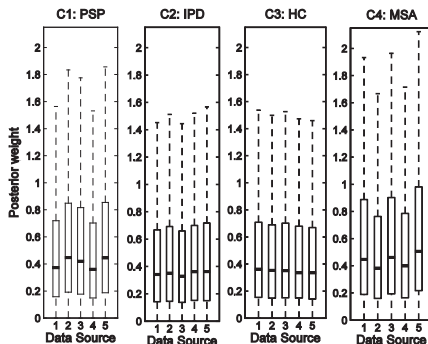
Multiclass classification with multiple regions

Analysis of brain regions

1. brainstem
2. bilateral cerebellum
3. bilateral caudate
4. bilateral middle occipital gyrus
5. bilateral putamen
6. all other regions

Method	Accuracy
GP classifier	0.614
SimpleMKL	0.229
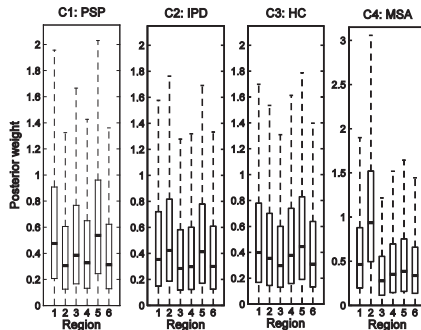
Brain Maps
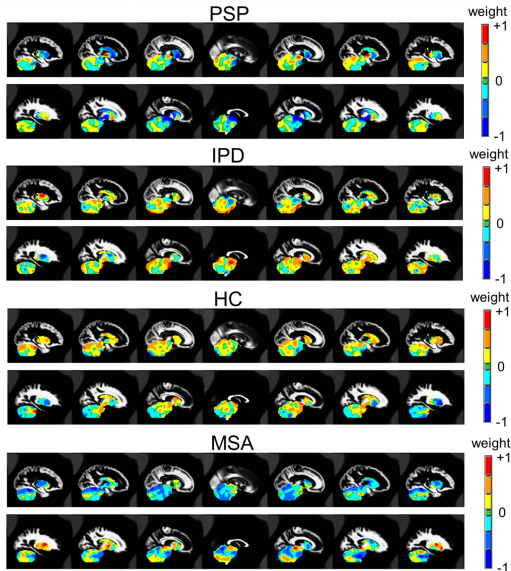
- Gaussian Processes yield flexible and interpretable nonparametric models

- Gaussian Processes yield flexible and interpretable nonparametric models
- Bayesian inference to accurately quantifying uncertainty in such models

- Gaussian Processes yield flexible and interpretable nonparametric models
- Bayesian inference to accurately quantifying uncertainty in such models
- Pseudo-Marginal MCMC offers a practical way to carry out exact Bayesian computations

Conclusions and ongoing work

- Gaussian Processes yield flexible and interpretable nonparametric models
- Bayesian inference to accurately quantifying uncertainty in such models
- Pseudo-Marginal MCMC offers a practical way to carry out exact Bayesian computations
- How to make exact Bayesian computations for Gaussian Processes scalable?

Acknowledgements

- Dr Andre F. Marquand (King's College London)
- Prof Mark Girolami (University College London)
- Dr Guido Sanguinetti (University of Edinburgh)
- Dr Alessandro Vinciarelli (University of Glasgow)

[1] M. Filippone and M. Girolami. Pseudo-Marginal Bayesian inference for Gaussian processes, *IEEE Transactions on Pattern Analysis and Machine Intelligence, to appear.*

[2] M. Filippone. Bayesian inference for Gaussian process classifiers with annealing and pseudo-marginal MCMC, In *ICPR, 2014.*

[3] M. Filippone et al. Probabilistic prediction of neurological disorders with a statistical assessment of neuroimaging data modalities. *Annals of Applied Statistics, 6(4):1883-1905, 2012.*

[4] A. F. Marquand et al. Automated, high accuracy classification of Parkinsonian disorders: a pattern recognition approach. *PLoS ONE, 2013.*

[5] M. Filippone et al. A comparative evaluation of stochastic-based inference methods for Gaussian process models. *Machine Learning, 93(1):93-114, 2013.*