University of Glasgow

Bayesian Inference for Gaussian Process Classifiers with Annealing and Pseudo-Marginal MCMC

Maurizio Filippone

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

International Conference on Pattern Recognition
August 26th, 2014

Motivation - Beyond cross-validation

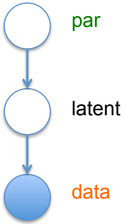
Probabilistic Kernel Machines

- Retain nonlinearity/flexibility of kernel machines
- Handle to an "objective function" - the log-likelihood

$$\log[p(\mathrm{data}|\mathrm{par})]$$

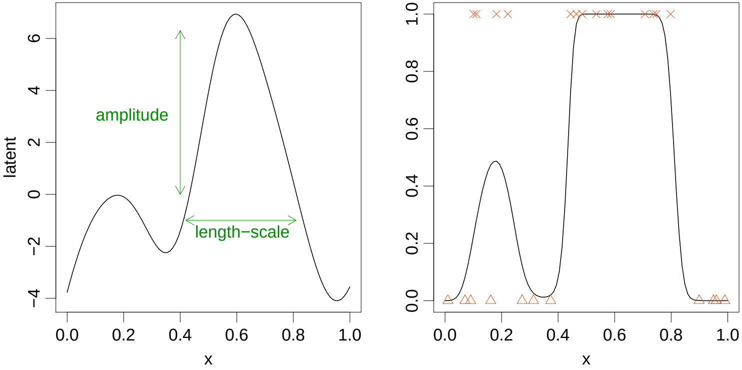
- It can be optimized wrt any number of parameters

- Graphical model



- Latent variables are assigned a Gaussian Process prior

Gaussian Process Models - Classification example



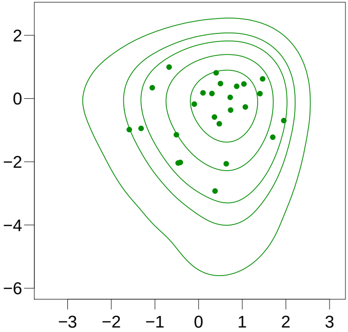
Motivation - Beyond optimization

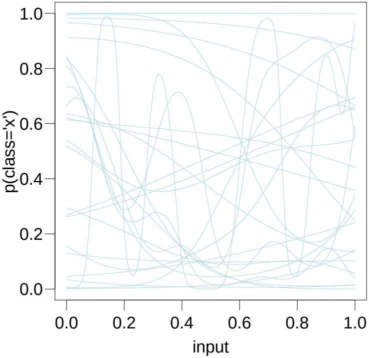
In some applications exact quantification of uncertainty is essential

- Optimization disregards any other "good" setting of kernel parameters
- Infer rather than optimize (Filippone and Girolami, IEEE TPAMI, 2014), (O'Harney et al., ICPR 2014)

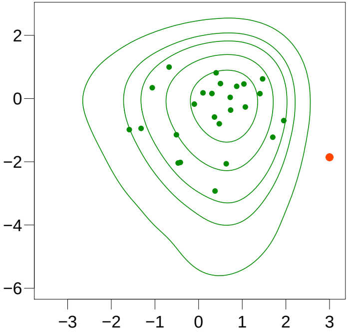
$$p(\mathrm{par}|\mathrm{data}) = \frac{p(\mathrm{data}|\mathrm{par})p(\mathrm{par})}{\int p(\mathrm{data}|\mathrm{par})p(\mathrm{par})d\mathrm{par}}$$

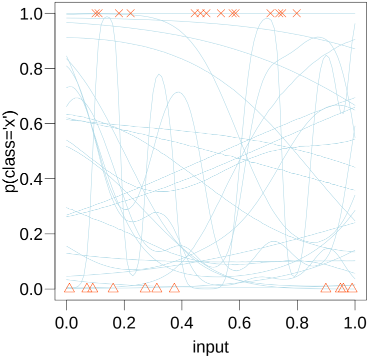
Bayesian Inference - Prior

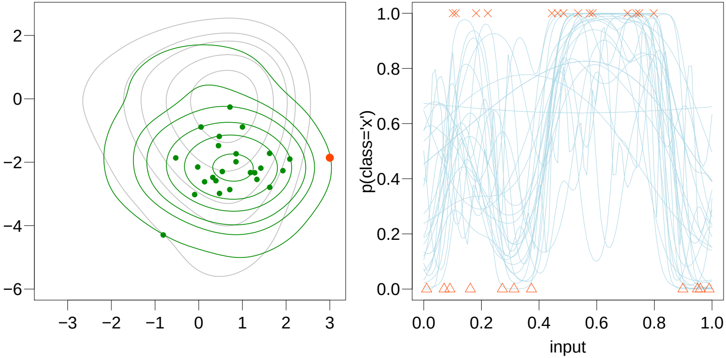




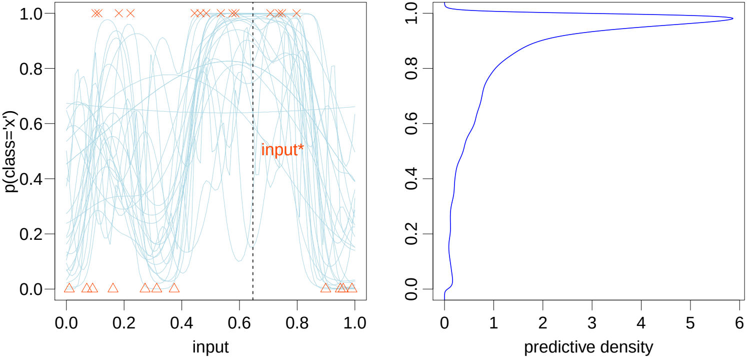
Bayesian Inference - Data



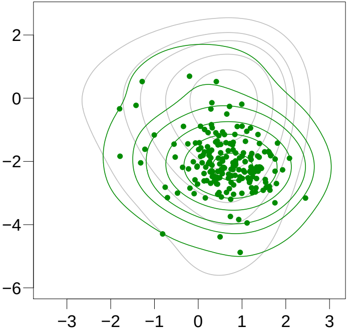



2
0
−2
−4
−6
−3
−2
−1
0
1
2
3
1.0
0.8
0.6
0.4
0.2
0.0
p(class='x')
0.0
0.2
0.4
0.6
0.8
1.0
input

Bayesian Inference and Predictions



- Draw samples according to the posterior density



Markov chain Monte Carlo (MCMC)

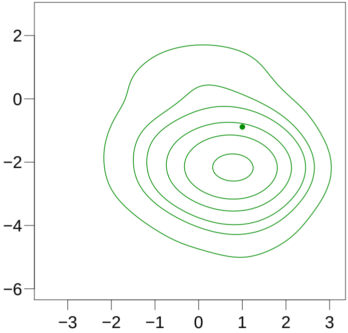
- Bayesian inference

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

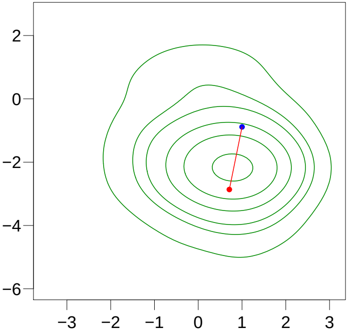
- Random walk sampler - accept a proposal with probability

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

- Explore the parameter space according to the density

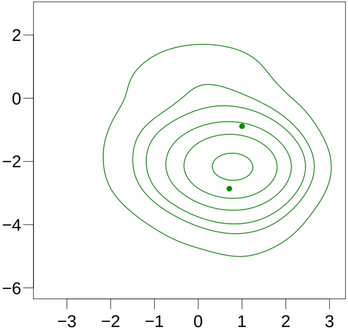


- Explore the parameter space according to the density

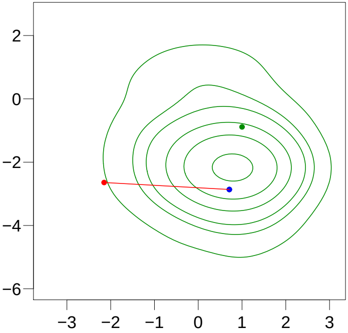


Markov chain Monte Carlo

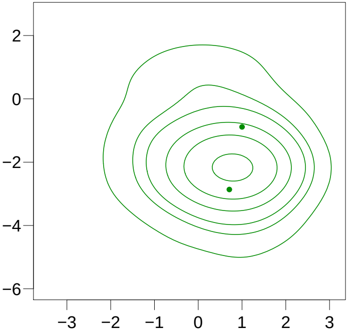
- Explore the parameter space according to the density



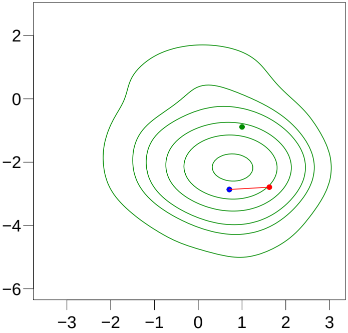
- Explore the parameter space according to the density



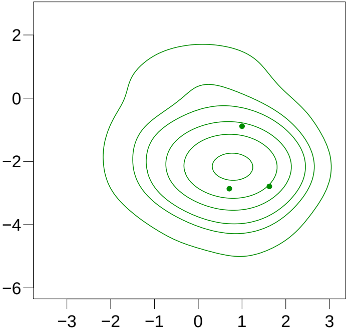
- Explore the parameter space according to the density



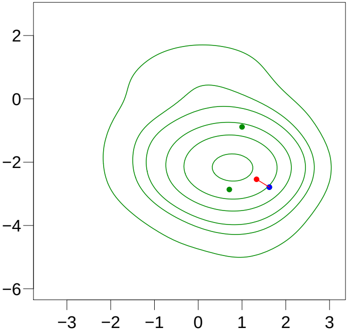
- Explore the parameter space according to the density



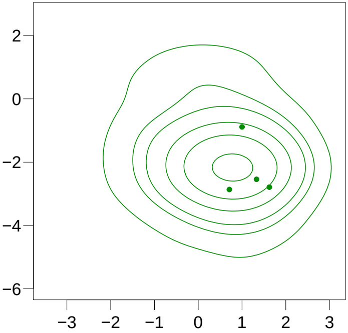
- Explore the parameter space according to the density



- Explore the parameter space according to the density

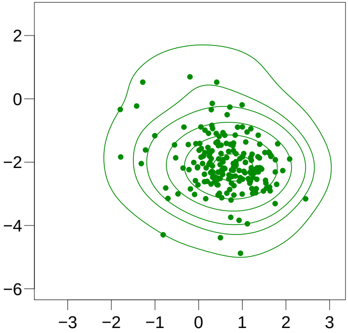


- Explore the parameter space according to the density



Markov chain Monte Carlo

- Explore the parameter space according to the density

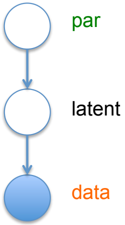


How can we draw from $p(\mathrm{par}|\mathrm{data})$ in these models?

- Marginal likelihood

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

is unavailable analytically



Pseudo-Marginal MCMC

- Replacing posterior by an unbiased estimate

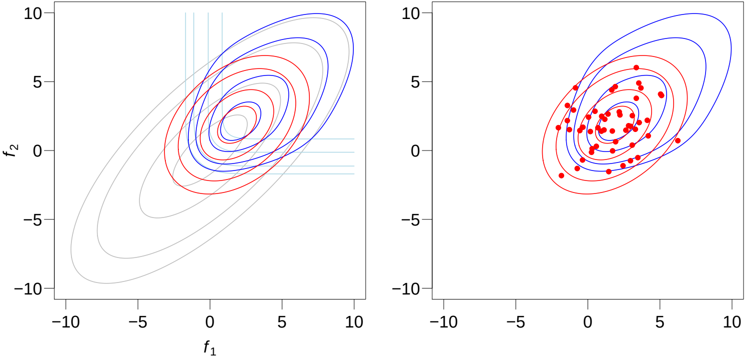
$$\min\left(1, \frac{\tilde{p}(\mathrm{par}'|\mathrm{data})}{\tilde{p}(\mathrm{par}|\mathrm{data})}\right)$$

retains correctness of the MCMC approach (Andrieu and Roberts, AoS, 2009), (Filippone and Girolami, IEEE TPAMI, 2014)

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

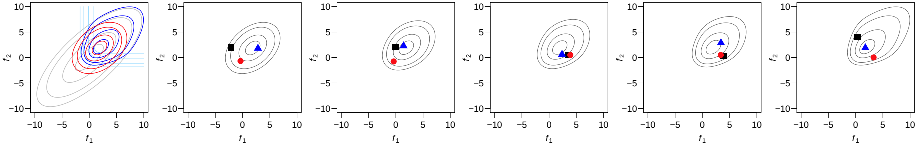
Importance Sampling estimator

- Approximate posterior over latent variables
- Then estimate $p(\text{data}|\text{par})$ using importance sampling



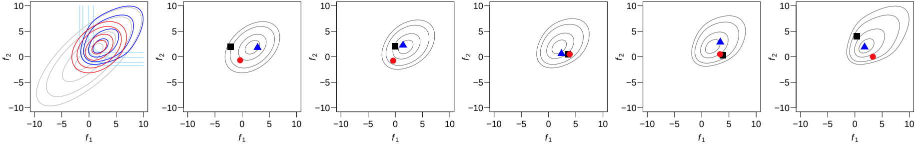
Annealed Importance Sampling (Neal, S&C, 2001)

- Annealing from an approximation

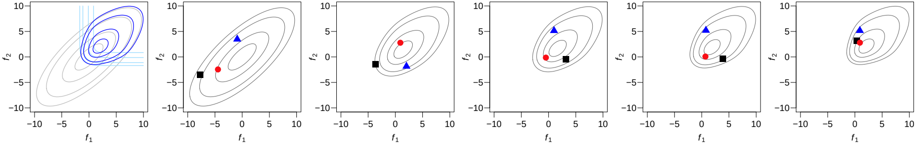


Annealed Importance Sampling (Neal, S&C, 2001)

- Annealing from an approximation



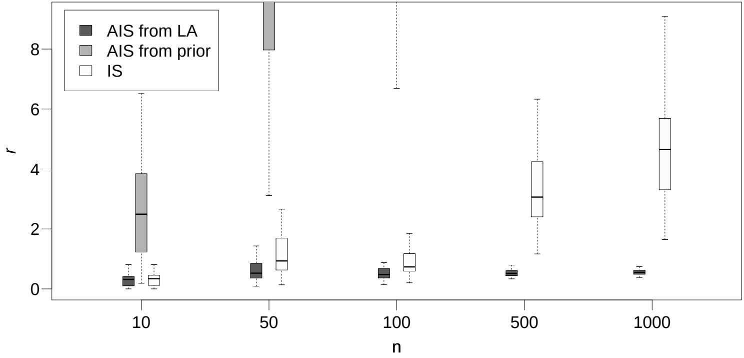
- Annealing from the prior



Comparison between AIS with IS

Analysis of the variance of the AIS and IS estimators

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



Acceptance rate MCMC

RBF Kernel

N_{imp}	Glass $n = 214,\ d = 9$		Thyroid $n = 215,\ d = 5$		Breast $n = 682,\ d = 9$	
	IS	AIS	IS	AIS	IS	AIS
1	2.8(1.6)	5.2(1.9)	1.1(1.0)	3.2(2.3)	17.9(2.4)	28.0(2.7)
10	10.4(3.1)	11.4(5.3)	4.1(3.8)	6.4(3.9)	30.5(4.1)	36.4(3.5)

RBF Kernel

N_{imp}	Pima $n = 768,\ d = 8$		Banknote $n = 1372,\ d = 4$		USPS $n = 1540,\ d = 256$	
	IS	AIS	IS	AIS	IS	AIS
1	24.8(1.4)	29.3(2.6)	1.1(0.6)	3.2(3.9)	0.6(0.6)	1.1(1.0)
10	30.8(2.6)	30.8(1.7)	4.7(1.0)	12.6(4.1)	0.6(0.5)	2.0(0.4)

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 (Radbout University)
- Prof Mark Girolami (University of Warwick)
- 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.*