

After AAAI submission

Thursday, August 29, 2024 10:01 AM

In "Hierarchical Bayesian modelling of gene expression time series across irregularly sampled replicates and clusters",
<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-252>

Hierarchical Bayesian modelling of gene expression time series across irregularly sampled replicates and clusters - BMC Bioinformatics

Background Time course data from microarrays and high-throughput sequencing experiments require simple, computationally efficient and powerful statistical models to extract meaningful biological si...

"irregularly sampled replicates and clusters" is something what have been also talking about!

1. We may have to test on the "benchmark" dataset soon: "substantially improved performance over eight popular multiple output (multi-task) Gaussian process models and three multivariate volatility models on benchmark datasets, including a 1000 dimensional gene expression dataset", in <https://arxiv.org/pdf/1110.4411>
 1. also "These are often called 'multi-task' learning or 'multiple output' regression models. Capturing correlations between outputs (response variables) can be used to make better predictions"
2. my attempt to formulate "non-independent" latent GPs is, likely, done in 2002: "Coregionalization by linear combination of nonorthogonal components. ' Mathematical Geology, 34 (4):405–419, 2002",
<https://link.springer.com/article/10.1023/A:1015078911063>

OMG, I see my equation(s)! The four summations, which is (much) nicely explained in the author of the slide deck and my GP go to person N. Lawrence <https://www.jmlr.org/papers/volume12/alfarez11a/alfarez11a.pdf>

We knew long ago that we are not "novel in algorithms", but more toward "novel in integrated view and applications". For that matter, the latent map and visualization may need to be around

I suggest you also (quickly read) the 2011 paper above. It referred to Bonilla as "The intrinsic coregionalization model has been employed in

Bonilla et al. (2008) for multitask learning", which is the simpler version of LMC, but employed Cholesky decomposition, as we do, for numerical stability.

--> We have another "integration"!

my mindset expanded to implementation as we are covering HGP. On slide 5 of our shared research plan, I have (more than) the following:

- I.LMC and HGP implementation
- 1.GPy? <https://github.com/SheffieldML/GPy/tree/deploy>,
 - a.N. Lawrence's 2013 paper
 - [api] https://gpy.readthedocs.io/en/deploy/GPy.kern.src.html#GPy.kern.src.independent_outputs.Hierarchical
 - [Notebook] <https://notebook.community/SheffieldML/notebook/compbio/Hierarchical>
 - b.LMC, two notebooks in "**Multiple Output Gaussian Processes**", on webpage
<https://nbviewer.org/github/SheffieldML/notebook/blob/master/GPy/index.ipynb>