Gaussian Processes

Sebastian Liem
@sebastianliem

APLS 2017, Kasteel Woerden

GRAPPA
GRavitation AstroParticle Physics Amsterdam

Global Analysis - Statistical Inference

1. Pick point θ in model M
2. Predict $y = f(\theta) + \varepsilon$
3. Compare y with data D

SLOW for the LHC!

Repeat (cleverly) until we understand the impact of D on M.

Eliminating Bottlenecks with Machine Learning

Problem: If calculating predictions is expensive, inference becomes prohibitive.

E.g. Predicting signal events at the LHC. $N_s = L\sigma\epsilon$

Solution: Replace expensive calculation with a cheap surrogate function.

Use machine learning to construct this surrogate.

Why Gaussian Processes?

Non-parametric: No need to assume a functional form.

!!! **Probabilistic:** Produces posteriors, i.e. estimates error.

Bayesian: Can specify prior on the type of functions.

What are Gaussian Processes?

Definition: A collection of random variables where any finite number of which have a joint Gaussian distribution.

The random variables represent the function value f(θ) at location θ.

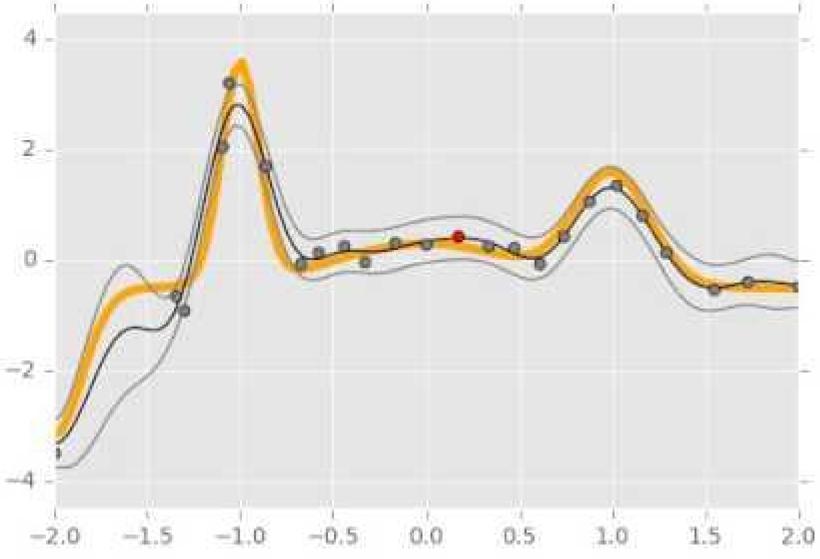
$$f(\theta) \sim \mathcal{GP}(m(\theta), k(\theta, \theta'))$$

Given some observations y = f(θ) + ε at θ (training data) we can predict

$$p(f(\theta') \mid \theta', y, \theta)$$

using normal Bayesian inference with Bayes' Theorem etc.

1D example



$$k(x, x') = A \exp\left[-\frac{1}{l}(x - x')^2\right] + B\delta(x - x')$$

Play at home!

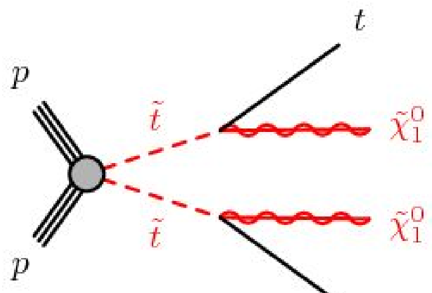
[Online demo, http://www.tmpl.fi/gp/](http://www.tmpl.fi/gp/)

https://www.youtube.com/watch?v=zTkdi-DVrXM

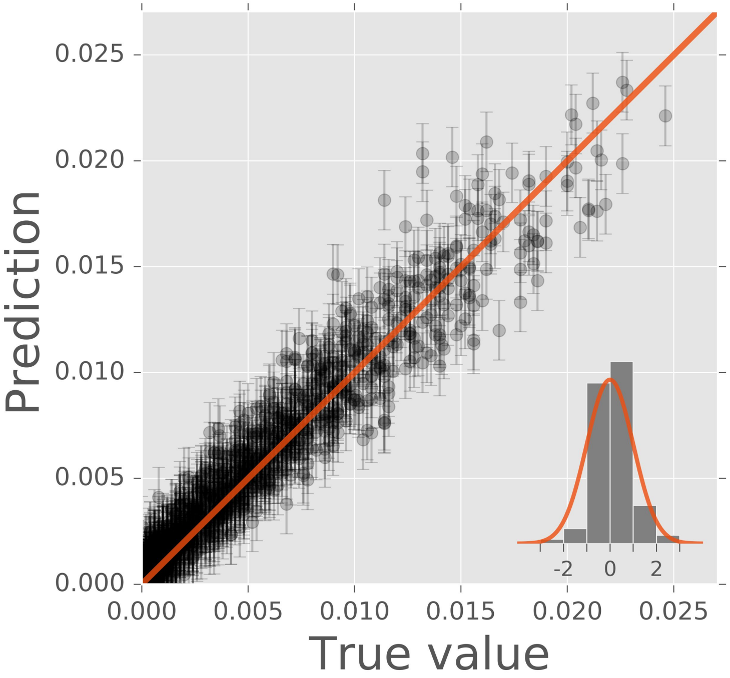
Natural SUSY (6 parameters)
18k training examples

Predicing the efficiency

$$N_s = L\sigma\epsilon$$



1-lepton signal region

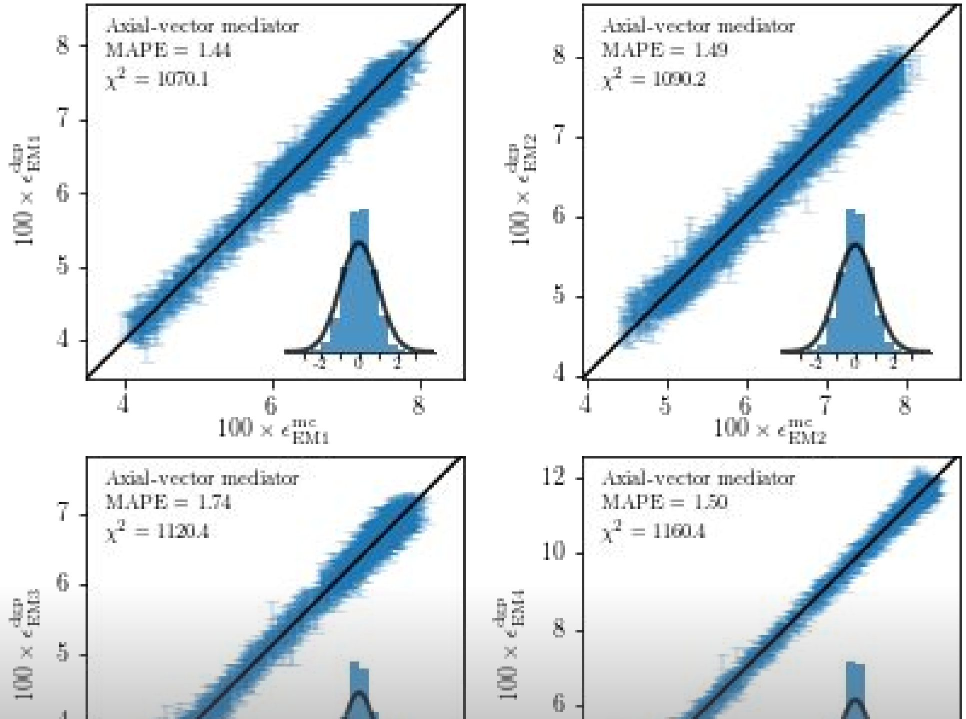


Accelerating the BSM interpretation of LHC data with machine learning

Gianfranco Bertone,[1] Marc Peter Deisenroth,[2] Jong Soo Kim,[3]
Sebastian Liem,[1] Roberto Ruiz de Austri,[4] and Max Welling[5]

arXiv:1611.02704

Simplified Model: Axial Vector Mediator ATLAS Monojet signal regions



Too many plots...

Current direction: Systematization

Systematize and semi-automate! Many things to accelerate.

Clever generation of the training data. Active learning.

Automate picking the best kernel structure. (arXiv:1302.4922)

(There is also interesting developments in how to train Bayesian Neural Networks.)

Additional slides

Gaussian Processes

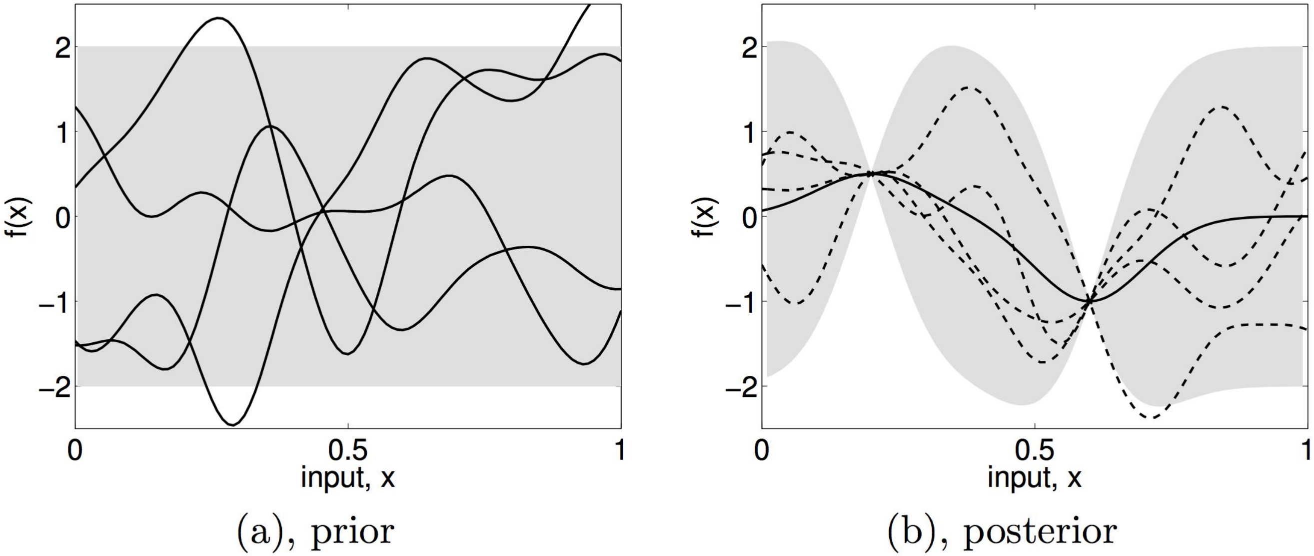


Figure 1.1: Panel (a) shows four samples drawn from the prior distribution. Panel (b) shows the situation after two datapoints have been observed. The mean prediction is shown as the solid line and four samples from the posterior are shown as dashed lines. In both plots the shaded region denotes twice the standard deviation at each input value x.

“Gaussian Processes for Machine Learning”
Rasmussen & Williams, 2006
www.gaussianprocess.org

Distributed Gaussian Processes

The standard Gaussian process scales badly with ***N*** the size of the training dataset. It involves inverting ***N*x*N*** matrices.

We use **distributed** Gaussian processes to avoid this. The data is randomly partitioned and on each partition a Gaussian process is defined. Predictions from each process is then combined.

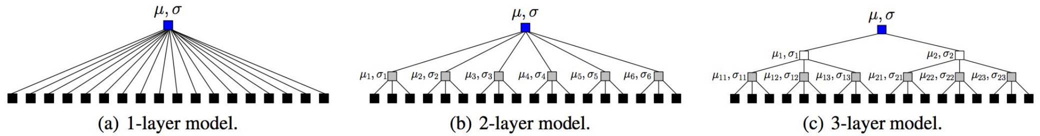


Figure 1. Computational graphs of hierarchical PoE models. Main computations are at the leaf nodes (GP experts, black). All other nodes recombine computations from their direct children. The top node (blue) computes the overall prediction.

"Distributed Gaussian Processes"
Deisenroth & Ng, arXiv:1502.02843

TensorFlow-based implementation:
https://github.com/ICL-SML/gptf

INPUT:
Natural SUSY

Natural because it stabilizes the electroweak scale without fine-tuning.

Only few SUSY states needs to be light.

$$\theta = \{\tan\beta,\ \mu,\ M_3,\ m_{\tilde{Q}_t},\ m_{\tilde{t}_R},\ A_t\}$$

Low-dimensional yet realistic theory.

We already had the training data from previous paper.

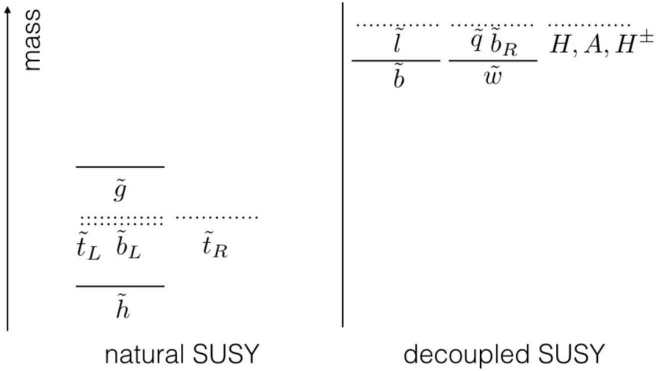
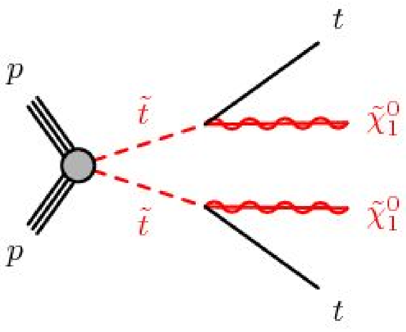


FIG. 1. The minimal natural SUSY mass spectrum on the left while the remaining supersymmetric particles are decoupled on the right.

“Natural SUSY: Now or Never?” Kim et al. arXiv:1606.06738

OUTPUT: Two Signal Regions



Defined in ATLAS-PHYS-PUB-2013-011

Looking direct stop production with HL-LHC, 14 TeV with 3000 fb^{-1}

Stops decay typically to top or b quarks, W/Z or Higgs bosons, and a LSP. Multiple jets, b-jets, large MET, possibly leptons.

1-lepton

MET > 750 GeV
m_T(lepton, MET) > 550 GeV

Total bkg: 21.1 ± 5.9

0-lepton

MET > 800 GeV
m_T(b-jet, MET) > 400 GeV

Total bkg: 12.2 ± 3.9

Training the Gaussian Processes

18647 models split into 16647 models for training and 2000 models for testing.

O(10 min) to train per signal region.

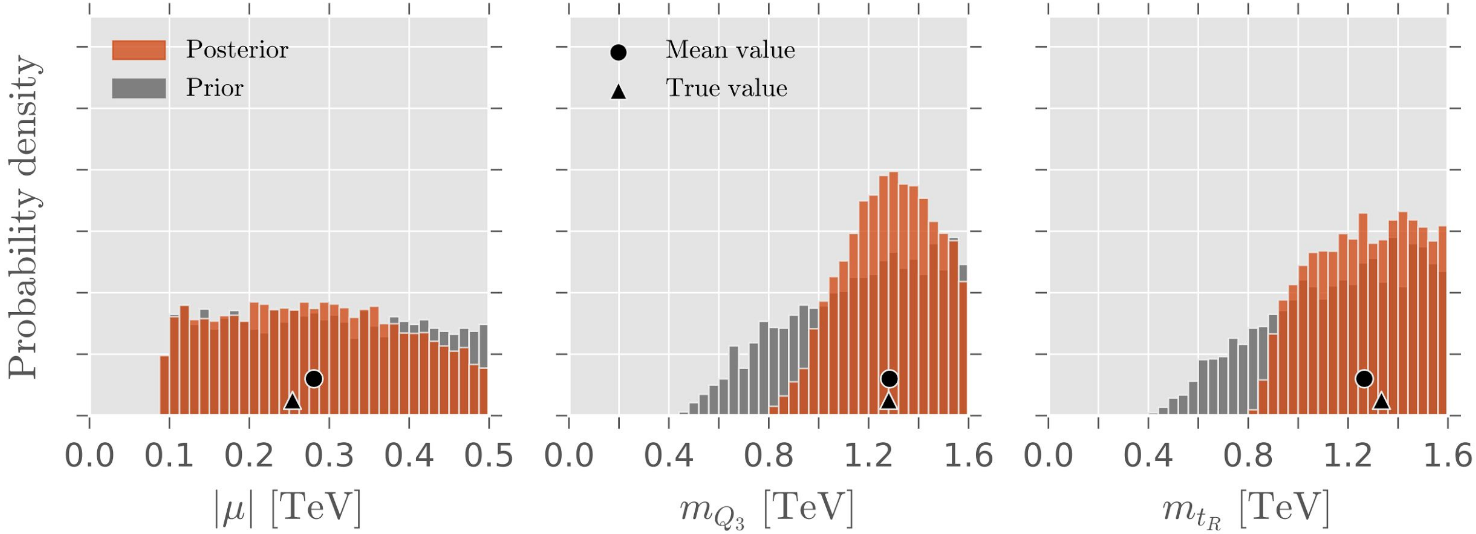
The lunch is not free, just cheaper!

SPheno, Pythia, NLLFAST, CheckMATE, Delphes etc. still needed to generate training data.

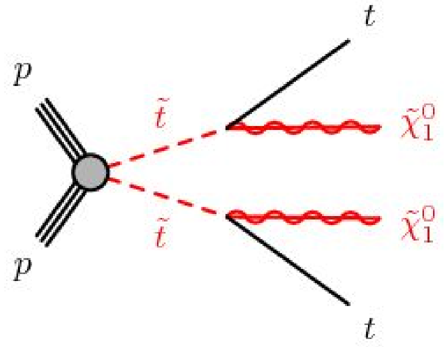
Models uniformly sampled from these ranges:

$$\begin{aligned} 0.1\,\mathrm{TeV} \leq\ & |\mu| \leq 1.0\,\mathrm{TeV}, \\ 0.1\,\mathrm{TeV} \leq\ & m_{\tilde{Q}_t} \leq 2.0\,\mathrm{TeV}, \\ 0.1\,\mathrm{TeV} \leq\ & m_{\tilde{t}_R} \leq 2.0\,\mathrm{TeV}, \\ 0.1\,\mathrm{TeV} \leq\ & |M_3| \leq 3.0\,\mathrm{TeV}, \\ & |A_t| \leq 3.0\,\mathrm{TeV}, \\ 1 \leq\ & \tan\beta \leq 20. \end{aligned}$$

All models avoid LEP-II chargino limit, and have reasonable Higgs boson mass.



Natural SUSY, mock signal at HL-LHC.
Stop production, 0-lepton and 1-lepton SR
MultiNEST ~109k likelihood eval
Simple example, 3D only.



Global analysis - Statistical Inference

Model M with parameters θ → Compute prediction $y = f(\theta) + \varepsilon$ → Compare with data D

Examples of M:
SUSY models
Simplified Model

dim $\theta \sim 4$-15

Examples of y:
Signal events in a bin
Relic density

Write down the likelihood.

Many different sources of data in dark matter, we do global analysis.

Sample (cleverly) until we understand the impact of D on θ.