Machine learning techniques for BSM phenomenology

Andy Buckley

Work with Alistair Shilton and Martin White

[arXiv:1106.4613]

IMPU-YITP MC Workshop, Kyoto, 2011-09-09

The problem

SUSY and other models have many parameters

Interpretation of experimental data fastest if unfolding of detector effects can be avoided

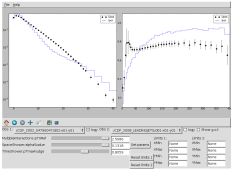
But detector simulation is *very* slow – 30 mins per event for ATLAS full sim!

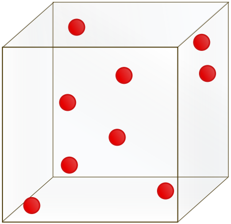
How to get coverage of the model space for constraints?

In general *many* params. But almost all pheno done in restricted models such as 5 param CMSSM. ATLAS and CMS constrain in m_0, $m_{1/2}$ only – definitely improveable!

Super-fast simulation for MC tuning

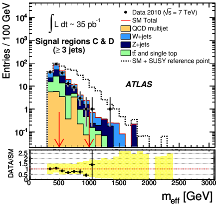
- We already deal with a similar problem in MC tuning!
- No full simulation, but again we have many parameters: MPI or hadronisation $\Rightarrow \mathcal{O}(5-10)$ params
- 1-10M events needed per energy/process to evaluate a param point. How to find the optimal point?
- Serial MCMC sampling intractable. PROFESSOR tuning system builds a polynomial *parameterisation* of every bin of every observable as a function of generator params
- Numerical optimisation is then trivial. It works!





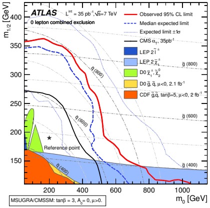
Parameterising an ATLAS SUSY analysis

Idea: use PROFESSOR to parameterise detector level cut-pass yields as a function of BSM model params.



Example: ATLAS' 2010 data 0-lepton analysis, signal region D [arXiv:1102.5290]:

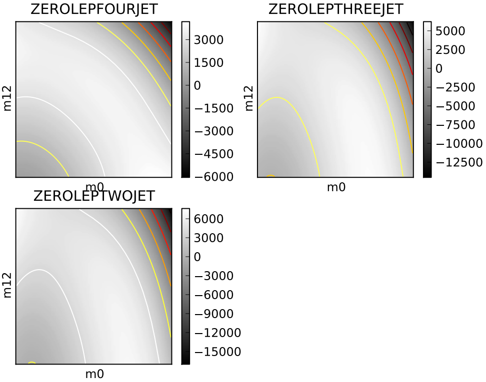
- $\geq$ 3 jets with $p_T > 40$ GeV and $p_T^{j1} > 120$ GeV,
- $E_T^{\rm miss} > 100$ GeV, $\Delta\phi(\text{jet}, p_T^{\rm miss})_{\rm min} > 0.4$,
- $m_{\rm eff} > 1000$ GeV, $E_T^{\rm miss}/m_{\rm eff} > 0.25$



ATLAS constrains m_0 & $m_{1/2}$ for $\tan\beta = 3$, $A_0 = 0$, $\mu > 0$

A first attempt…

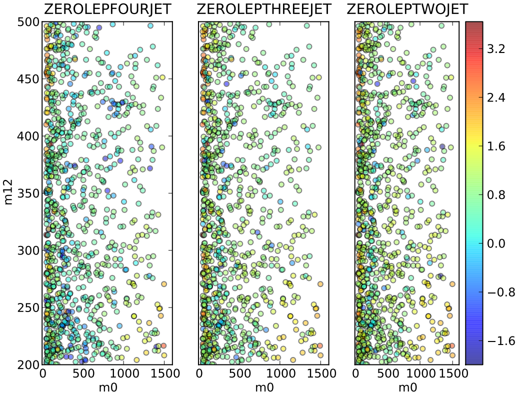
1000 HERWIG+AcerDet points from 4D CMSSM with $\mu > 0$: parameterise #events passing Region D cuts. A 1-bin observable!



Unphysical negative values on boundary!

A first attempt…

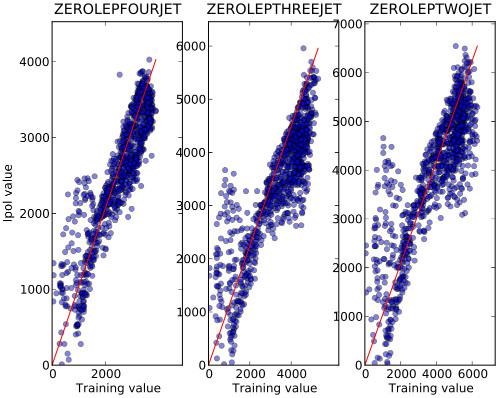
1000 HERWIG+AcerDet points from 4D CMSSM with $\mu > 0$: parameterise #events passing Region D cuts. A 1-bin observable!



Showing $\log_{10}$(% deviation): accuracy typically within 10%

A first attempt…

1000 HERWIG+AcerDet points from 4D CMSSM with $\mu > 0$: parameterise #events passing Region D cuts. A 1-bin observable!



Definite artefacts in MC / ipol correlation. **Not so good…**

Improving the parameterisation

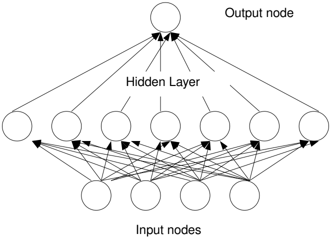
What went wrong? Problems particularly at the boundaries: fitted polynomials are not trustworthy near the edges of the space.

BUT – we don't need an algebraically constructed parameterisation since there is only one observable: fitting speed is not a serious issue. Thresholds and discontinuities in the param space.

Idea #2: use a parameterisation-free interpolating function:

- FBM Bayesian neural net
- SVMHeavy support vector machine

Training time for one bin $\sim$12 hrs!



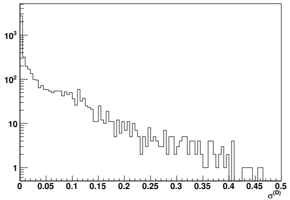
Sampling details

Sample from the 4D CMSSM with $\mu > 0$ again:

Parameter	Minimum	Maximum
$m_{1/2}$ [GeV]	50	1200
m_0 [GeV]	50	1200
A_0 [GeV]	-1000	1000
$\tan\beta$	2	60

20000 points each of 50k HERWIG+AcerDet events $\Rightarrow \sigma^{(D)}$

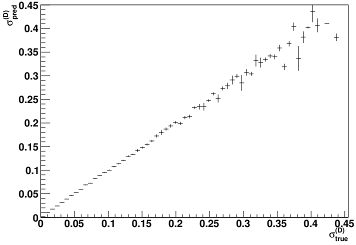
BNN/SVM will train to optimally fit the max *number* of training points. Implicit measure on the sampling space.
Flat sampling undersamples high-σ = low $m_{1/2}$ region. Sample $m_{0,1/2} \sim e^{-x/\lambda}$ with $\lambda = 200$ GeV.

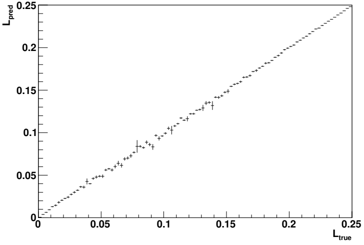


Validation

Predicted vs. true $\sigma^{(D)}$ and the Poisson likelihood L used to make the exclusion contour

BNN with 5k points; independent training and test sets



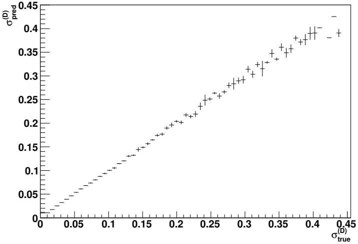


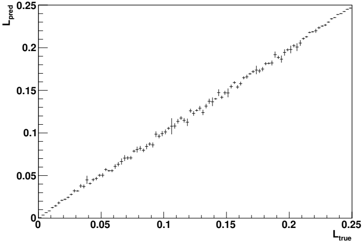
Excellent performance on both measures

Validation

Predicted vs. true $\sigma^{(D)}$ and the Poisson likelihood L used to make the exclusion contour

SVM with 5k points; independent training and test sets



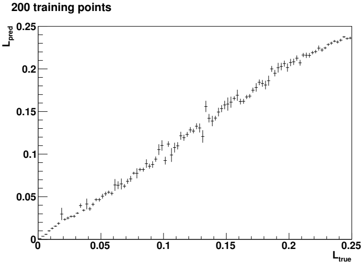


Slightly better at high $\sigma^{(D)}$ but overall slightly worse L

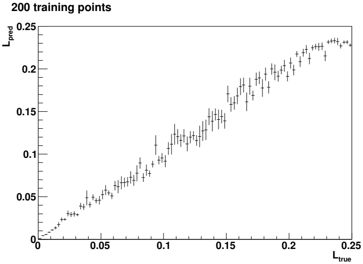
Likelihood convergence

200 training points

BNN



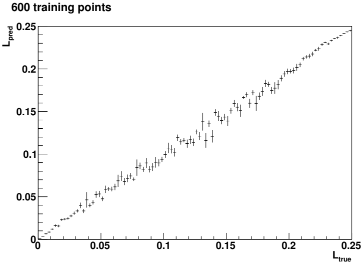
SVM



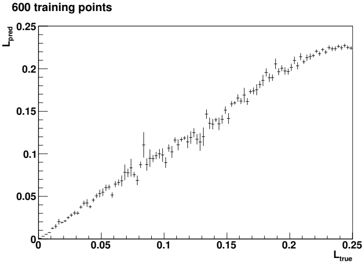
Likelihood convergence

600 training points

BNN



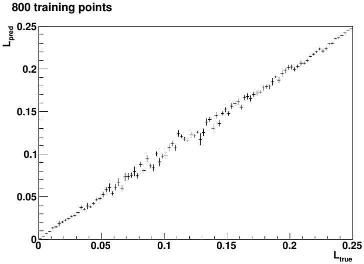
SVM



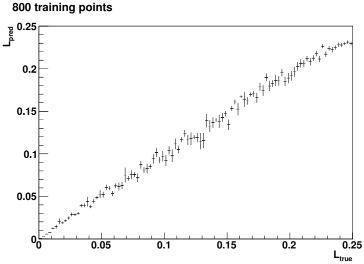
Likelihood convergence

800 training points

BNN



SVM

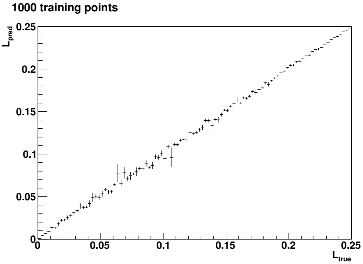


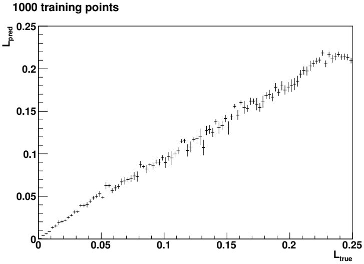
Likelihood convergence

1000 training points

BNN

SVM

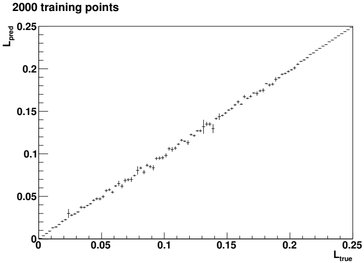




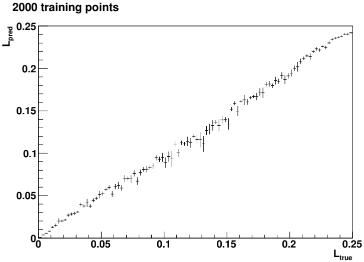
Likelihood convergence

2000 training points

BNN



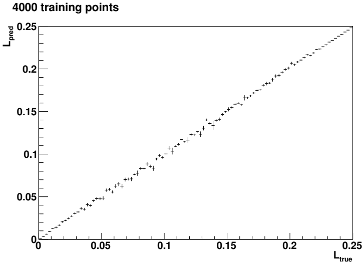
SVM



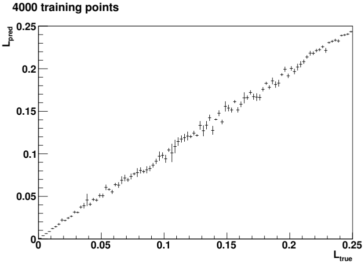
Likelihood convergence

4000 training points

BNN

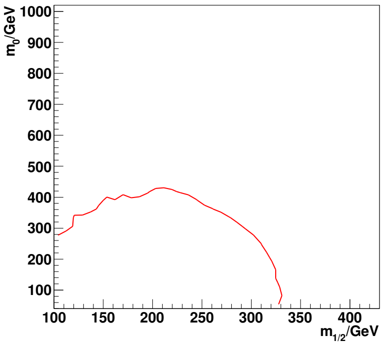


SVM

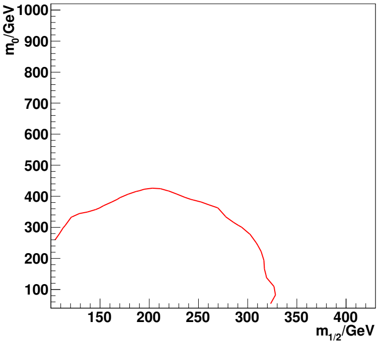


Exclusion contours from coarse grid scanning

True



BNN/SVM 5k



Compatible 95% exclusion results in m_0–$m_{1/2}$ plane for "ATLAS-like" fixed $\tan\beta = 3$, $A_0 = 0$.

Summary and outlook

- BSM model exclusion becomes exp. weaker as # params increases
- Particularly when detector (full or fast) simulation needed
- Parameterisation has been a big success in MC tuning: idea has been applied to an example BSM parameter space and analysis
- PROFESSOR polynomial ipols and SVD are crippled by parameterisation bias. For small numbers of observables, neural net or SVM approaches are much better
- Outlook:
  - extend to more complex signatures and models
  - use multiple observables in likelihood, handle systematics
  - use newer/better MC generators + NLO cross-sections
  - better detector simulation: AtlFast-2 in ATLAS, ??? outside

Thank you!

Many thanks to YITP, IMPU, Mihoko, Bryan, Rie. . . for the opportunity to spend a week in Kyoto, surrounded by QCD: