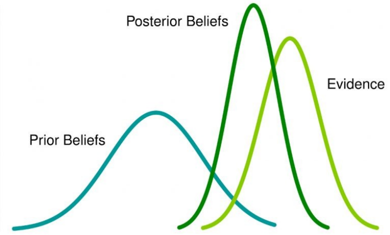
Bayesian Machine Learning in HEP - is it a good idea?

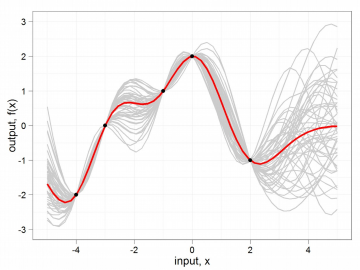
Vidhi Lalchand

Nov 30th, 2018

3 ideas in 10 mins

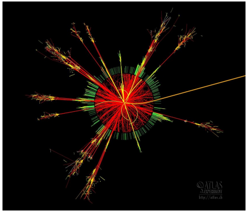
- Unpacking the term ‘Bayesian Machine Learning’





- Quantifying uncertainty in predictions

- Useful in HEP tasks like density estimation?



Conventional vs. Bayesian

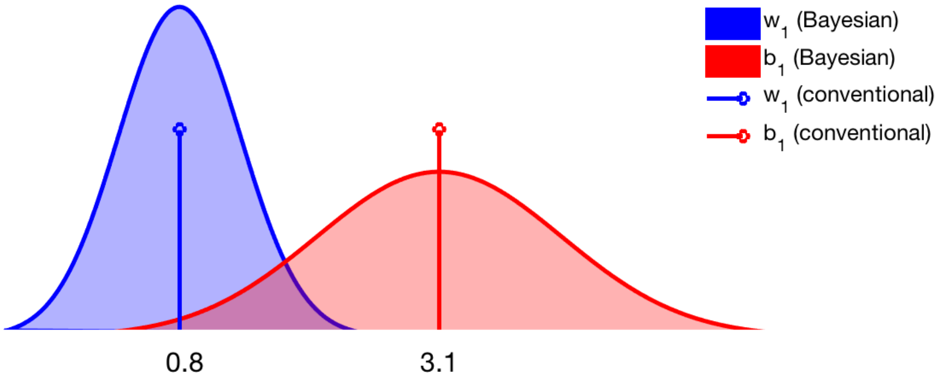
Conventional Machine learning:

- Step 1: Data $D = \{x_i, y_i\}_{i=1}^{N}$ + Parametric model, eg: $y_i = w_1 x_i + b_1$, $\theta = \{w_1, b_1\}$.
- Step 2: Train the model $\rightarrow$ **'learn'** the parameters $\theta \rightarrow \left\{\hat{w}_1, \hat{b}_1\right\}$.
- Step 3: Make predictions for unseen x_* by plugging in, $y_* = \hat{w}_1 x_* + \hat{b}_1$.

$$\boxed{\text{Training step: } \underset{\theta}{argmin}\ \mathcal{L}(\theta)}$$

Parameters θ are fixed, unknown quantities

Predictions: Point estimates vs. Distributions

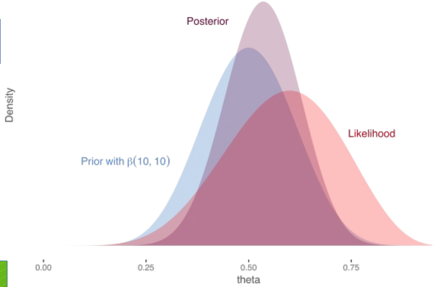


What is Bayesian Machine Learning?

Bayesian Machine learning:

$D = Data$
$\theta = parameters$

- Step 1: Select prior probability distribution $p(\theta)$ and the Data Likelihood $p(D|\theta)$
- Step 2: Bayesian inference step → "**learn**" the posterior distribution $p(\theta|D)$
- Step 3: Prediction step → derive predictive distribution for new input $p(x_*|D)$



Likelihood
Prior
Posterior

$$p(\theta|D) = \frac{p(D|\theta)\,p(\theta)}{p(D)}$$

Evidence

where,

$$p(D) = \int p(D|\theta)\,p(\theta)\,d\theta$$

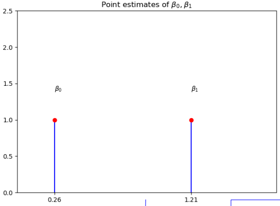
$$p(x_*|D) = \int p(x_*|D,\theta)\,p(\theta|D)\,d\theta$$

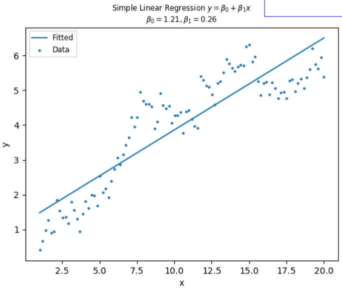
Bayes' rule

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

$$P(B) = \sum_i P(B|A)\,P(A_i)$$

Conventional



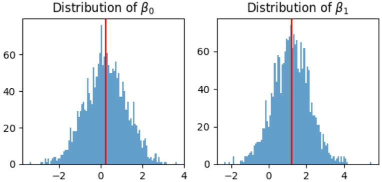


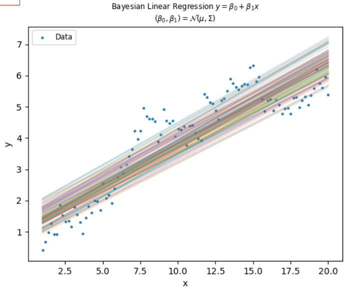
$$\theta = (\beta_0, \beta_1)$$

Fixed, unknown quantities

Random variables

Bayesian





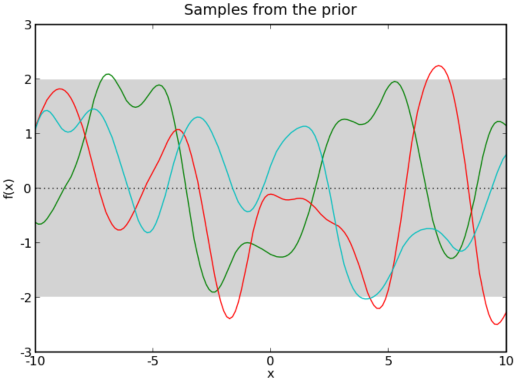
Informal

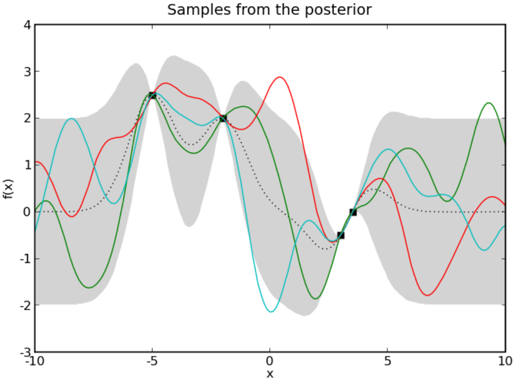
Bayesian machine learning is about using the calculus of probabilities to conduct inference on unknown quantities (.i.e. parameters) and making predictions using the inferred quantities with a careful representation of uncertainty.

- Non-Bayesian setting: Learning happens through optimising a loss function, this involves computing **derivatives**.
- Bayesian setting: Learning happens through the application of Bayesian inference, this involves **integration**.

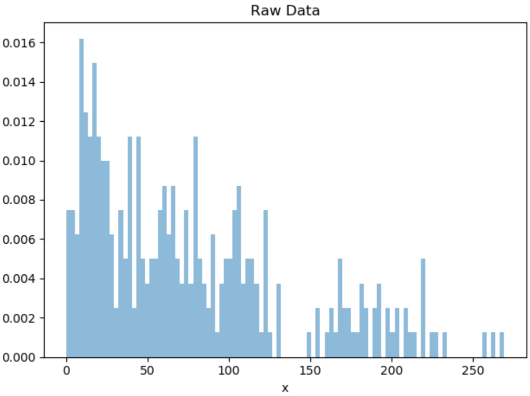
Bayesian method for modelling functions → Gaussian Processes

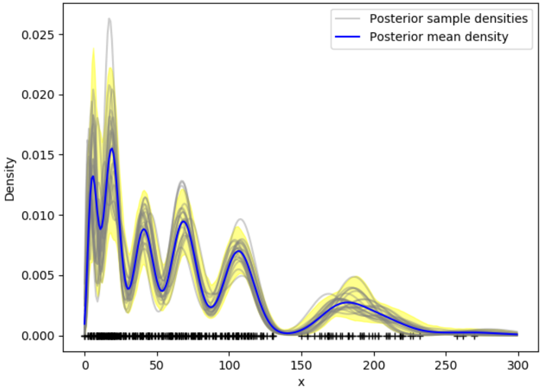
$$\theta = \mathbf{f}$$
$$p(\mathbf{f}) \rightarrow p(\mathbf{f}_*|\mathbf{f})$$
Prior → Posterior



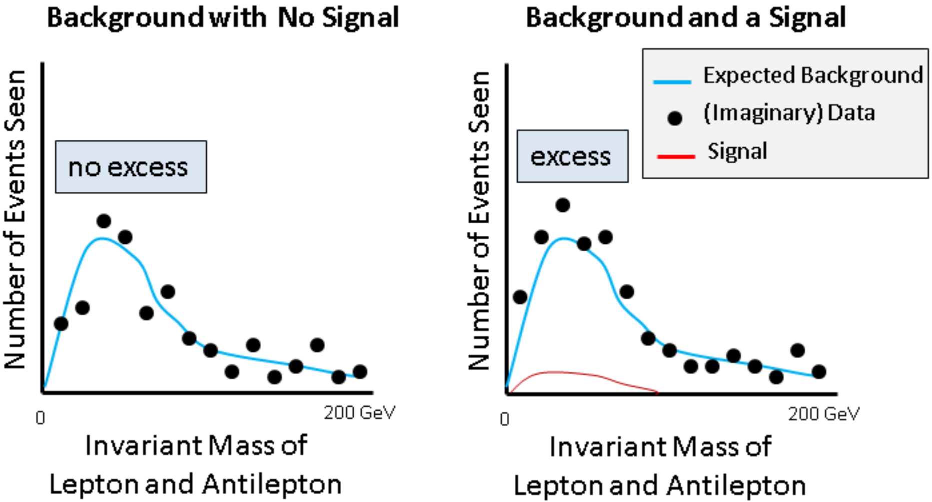


Bayesian method for modelling densities



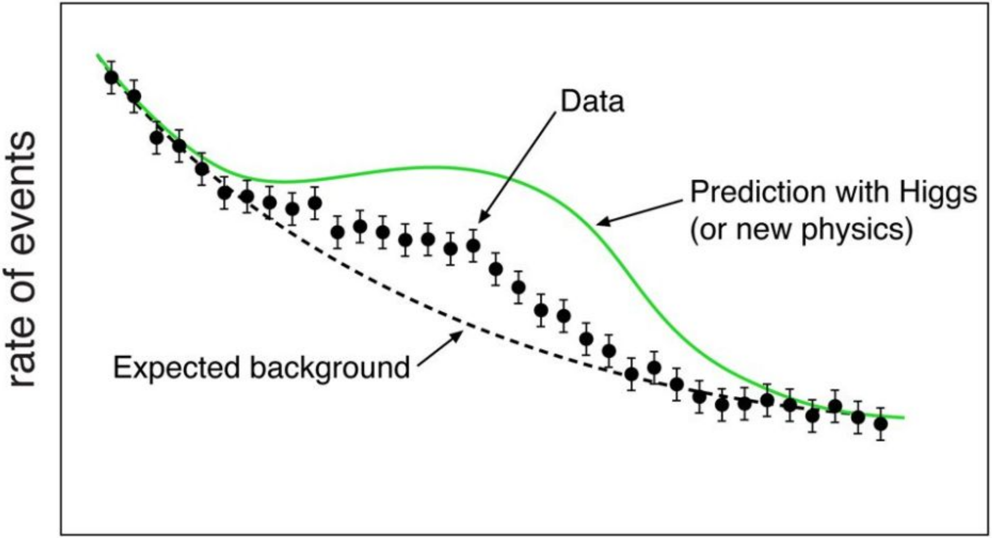


Density Estimation in HEP

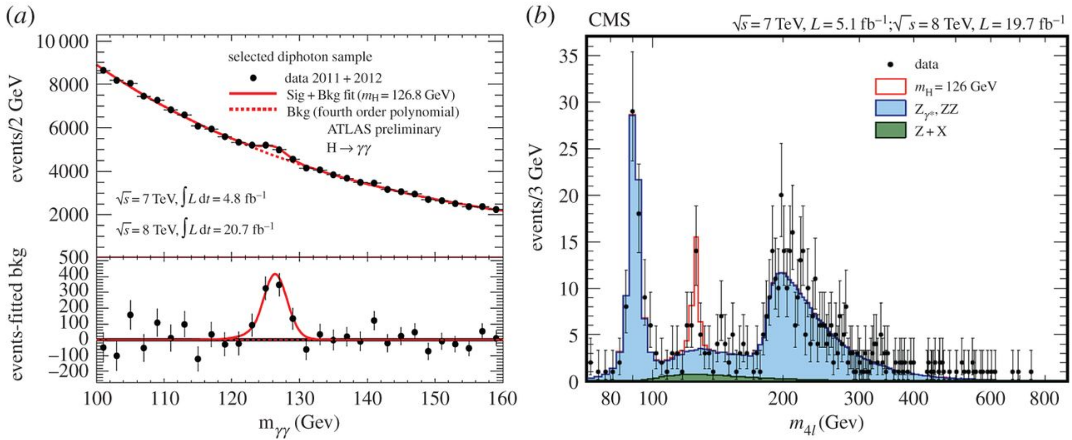


M. Strassler 2011

Density Estimation in HEP



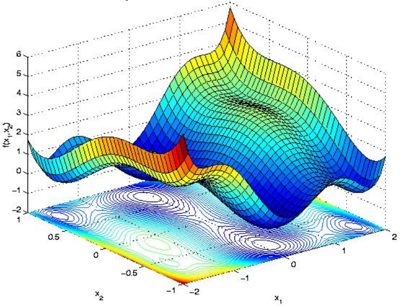
Density Estimation in HEP



Going Bayesian!.. is it a good idea?

Yes:

- Better understanding of the uncertainty around predictions.
- When dealing with complex, multi-modal and high dimensional data, it is much better to integrate rather than optimise.



(caveat)

Pain point: The integrals encountered in Bayesian inference are intractable and one has to resort to numerical techniques like Markov chain monte carlo and its variants.