Bayesian Neural Networks for GW parameter Estimation

Hongyu Shen, Wei Wei
Dept. of Electrical and Computer Engineering
Dept. of Physics
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign

Spotlights

- Hands-on Linear Regression Model

- Bayesian Neural Network on GW Parameter Estimation

Backgrounds

- Assume Likelihood function on noisy observations x:

$$p(x \mid m_1, m_2)$$

- Assume flat prior on parameters (e.g. masses m_1, m_2)

$$p(m_1, m_2)$$

- Apply Bayesian rules to achieve the the posterior of the parameters

$$p(m_1, m_2 \mid x) \propto p(x \mid m_1, m_2) p(m_1, m_2)$$

Backgrounds

- Posterior function is not easy to evaluate due to difficulties in exploring the whole space of the parameters (red and blue)-> use MCMC sampling instead -> time-consuming

$$p(m_1, m_2 | x) = \frac{p(x | m_1, m_2)p(m_1, m_2)}{p(x)}$$

- We can make improvements with Bayesian Neural Networks!

Content

- Hands-on Linear Regression Model

- Apply ideas from Linear Regression on GW parameter Estimation

Resource

https://github.com/skyve2012/BayesianNetTutorial

Linear Regression

- Consider the following problem:

$$y = ax + b + \epsilon$$

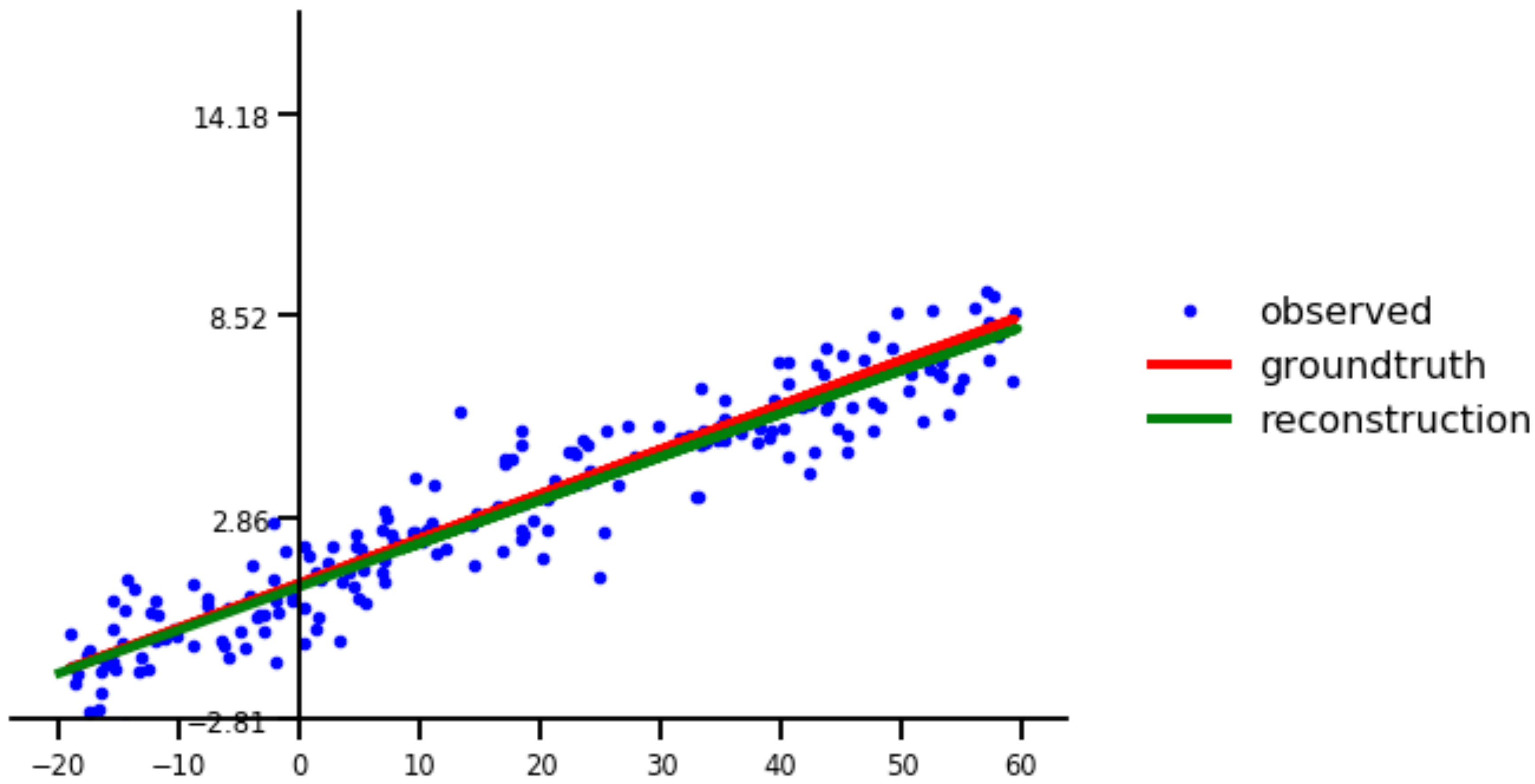
- y and x are data; a and b are parameters to be learned. ϵ is universal noise across all values of x (0 mean, fixed variance)

- Best solution is least square under the linear assumption in the true model.

$$\min_{a,b} ||ax + b - y||^2 = max_{a,b} p(y|x; a, b)$$

$$p(y|x; a, b) \sim N(ax + b, \sigma_\epsilon^2)$$

Example


14.18
8.52
2.86
−2.81
−20
−10
0
10
20
30
40
50
60
observed
groundtruth
reconstruction

Linear Regression

- What if the problem is slightly changed:

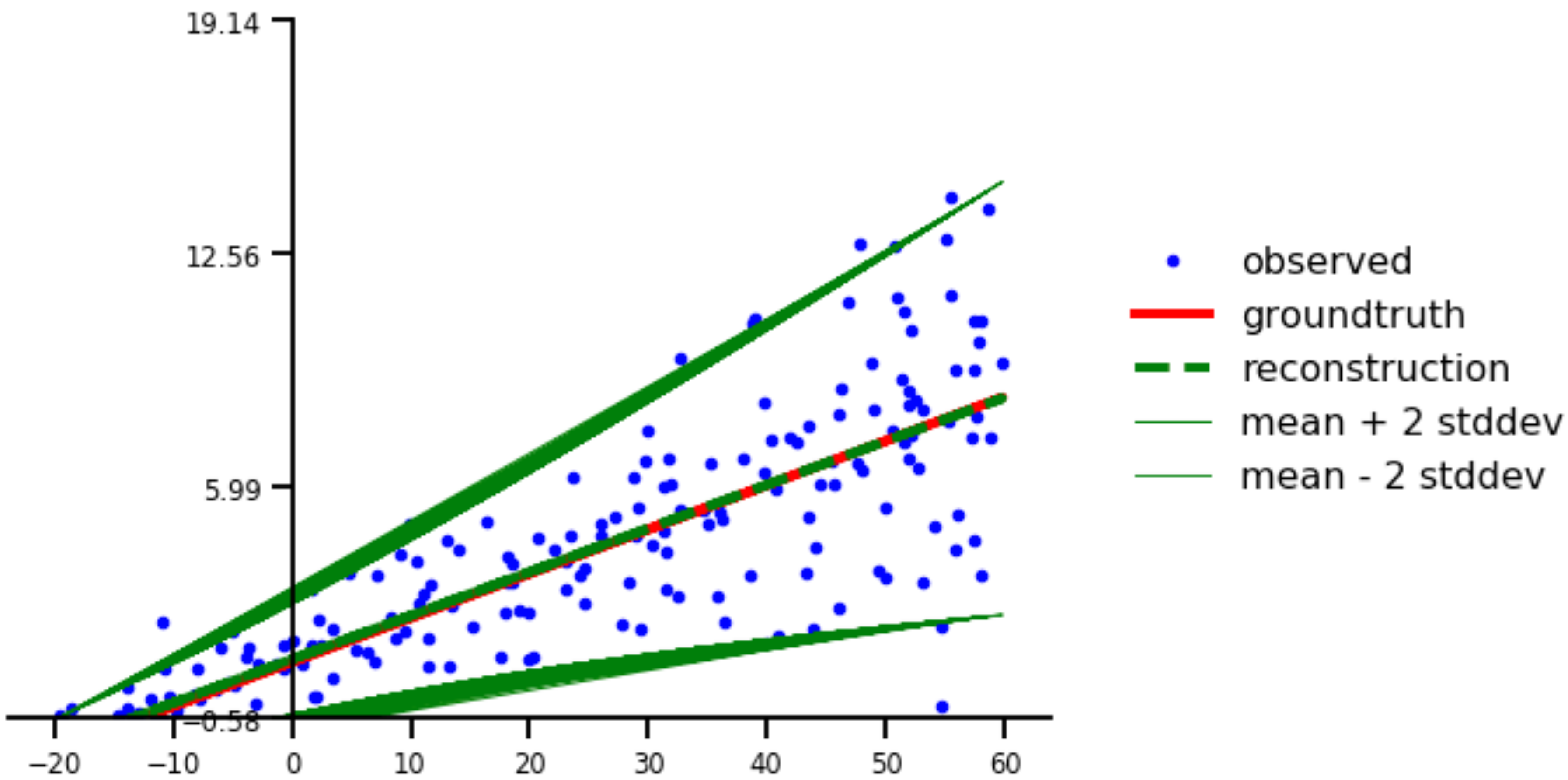
$$y = ax + b + \epsilon(x)$$

- ϵ is dependent on x in this case (0 mean).
- One way is to use weighted least square
- The other way is to treat $\epsilon(x) = \epsilon_\theta(x)$

$$max_{a,b,\theta} p(y \mid x; a, b, \theta)$$

$$p(y \mid x; a, b, \theta) \sim N(ax + b, \sigma^2_{\epsilon_\theta(x)})$$

Example



19.14
12.56
5.99
−0.58
−20
−10
0
10
20
30
40
50
60
observed
groundtruth
reconstruction
mean + 2 stddev
mean - 2 stddev

Linear Regression

- What if we want to also consider the variance on the estimation $\hat{a}, \hat{b}$ given the distribution of the data: $p(x, y)$. Or equivalently, $p(a, b \mid x)$.

- Consider the previous problem: $max_{a,b,\theta} p(y \mid x; a, b, \theta)$

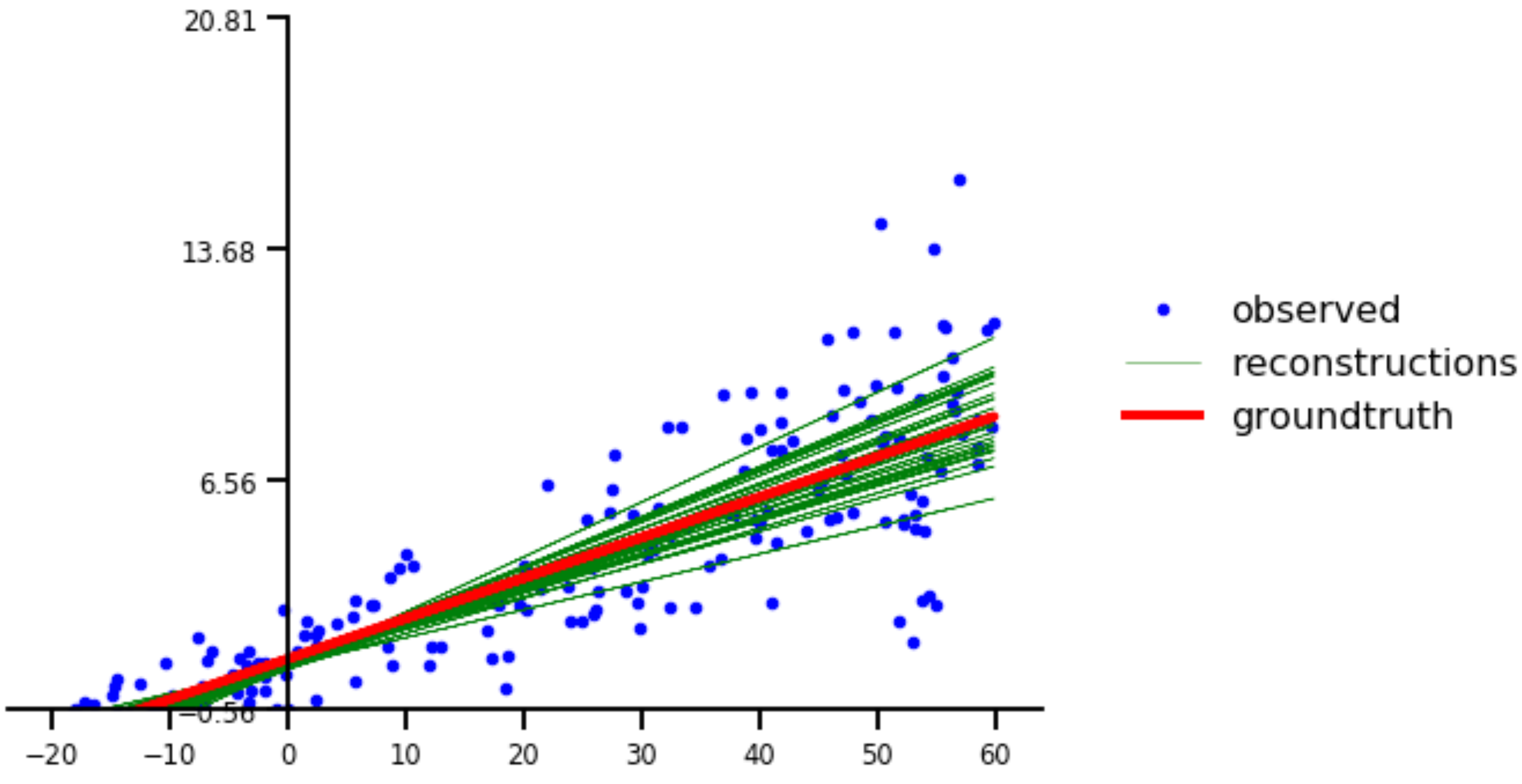
- We re-write the problem as: $max_{\theta,w} p(y \mid x; \theta, w)$

$$p(y \mid x; \theta, w) = \int_{a,b} p(y \mid a, b, x; \theta) p_w(a, b \mid x) dadb$$

- The blue part is can be learned via Variational Inference

https://github.com/skyve2012/BayesianNetTutorial

Example



20.81
13.68
6.56
−0.56
−20
−10
0
10
20
30
40
50
60
observed
reconstructions
groundtruth

Linear Regression

Demo

GW Parameter Estimation

- This is no difference to the linear regression
- Linear Regression: $y = ax + b + \epsilon(x)$
- GW Parameter Estimation: $y = f(x) + \epsilon(x)$, $y = f(x, \epsilon(x))$
- Assume we have a parameterized model $\hat{f}_\tau(x)$:

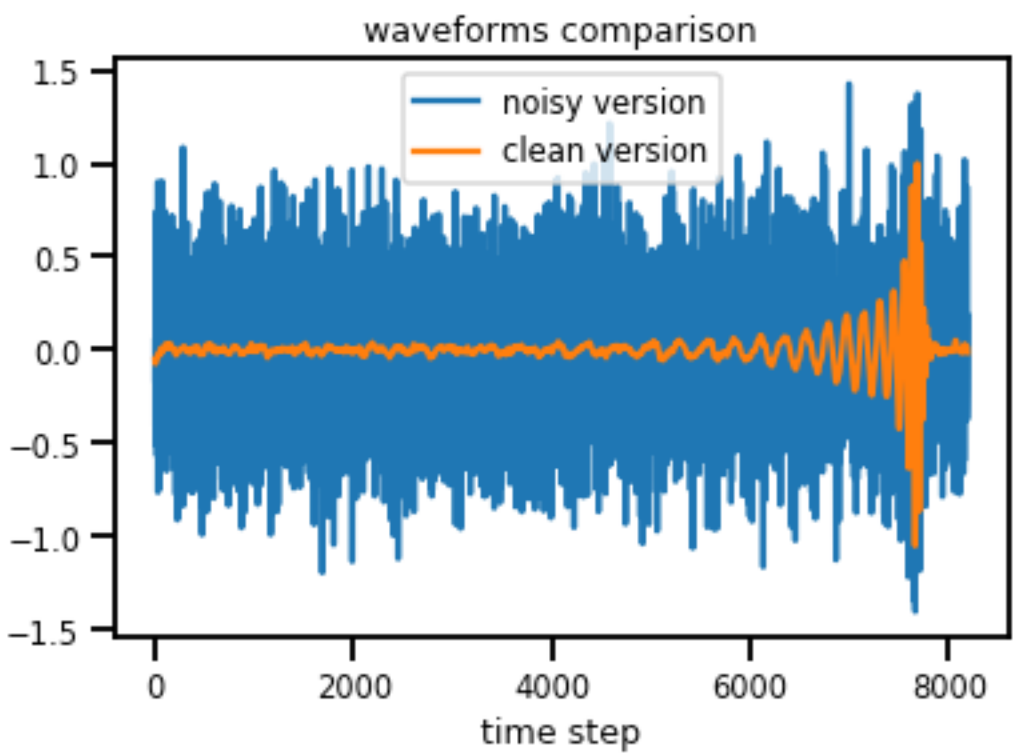
$$max_{\theta,w} p(y \mid x; \theta, w)$$

$$p(y \mid x; \theta, w) = \int_\tau p(y \mid \tau, x; \theta) p_w(\tau \mid x) d\tau$$

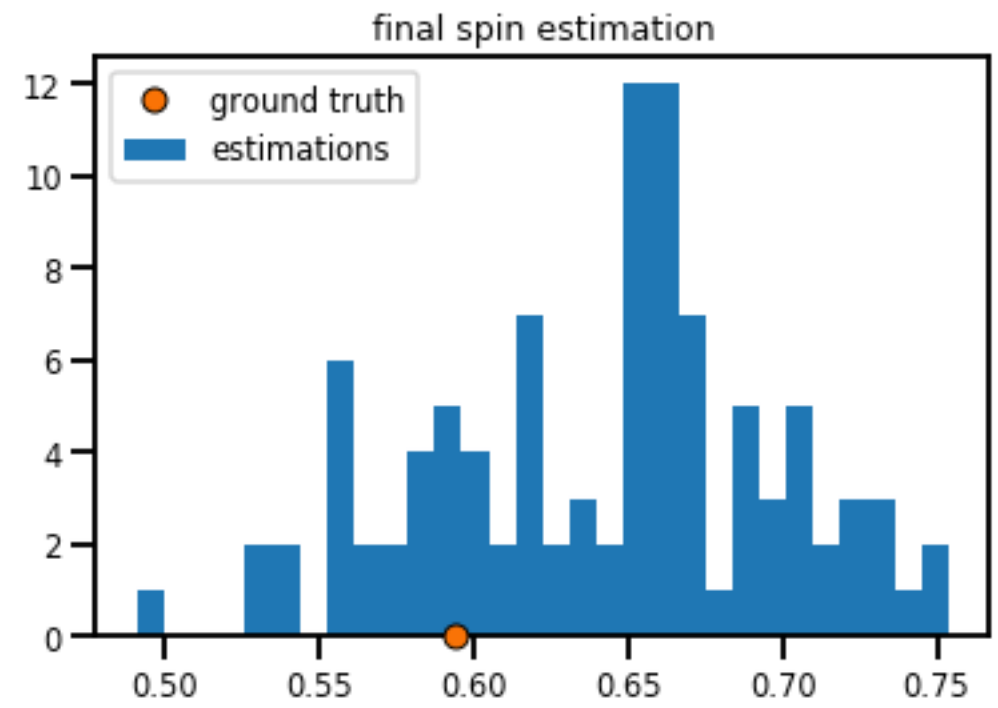
https://github.com/skyve2012/BayesianNetTutorial

Example

Noisy Waveform



Estimated Final Spin



GW Parameter Estimation

Demo