Bayesian Machine Learning

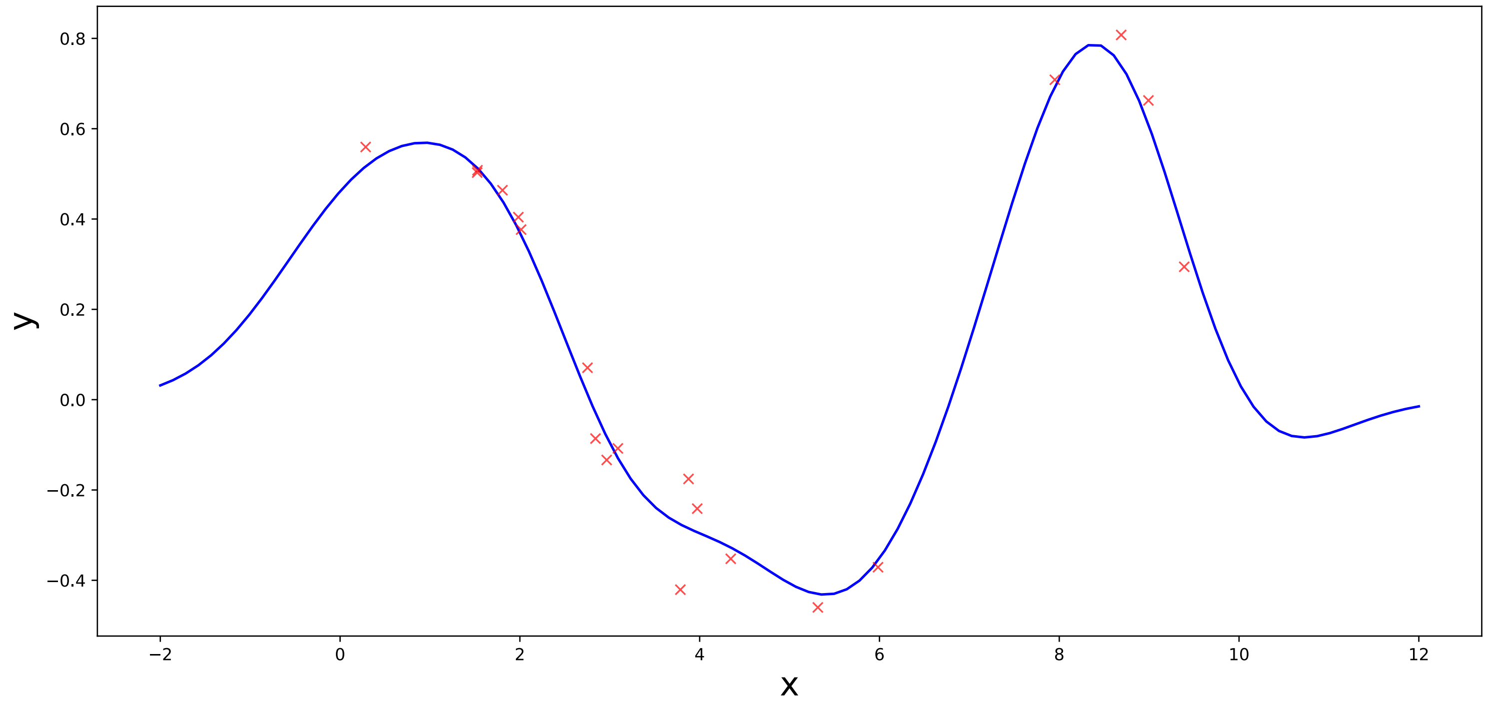
Rajbir-Singh Nirwan

January 24, 2018

Agenda

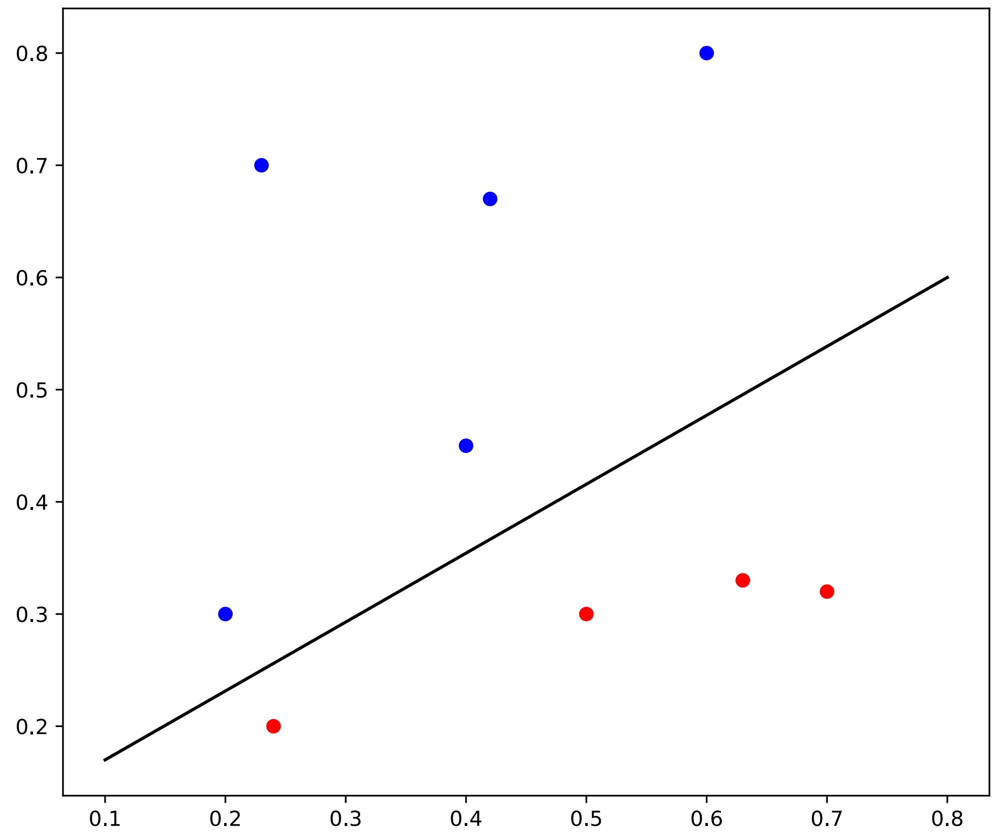
- Overview - Problems and Solutions
- Deep Neural Networks
- Probabilistic Modeling
- Hands-On Exercises

Regression



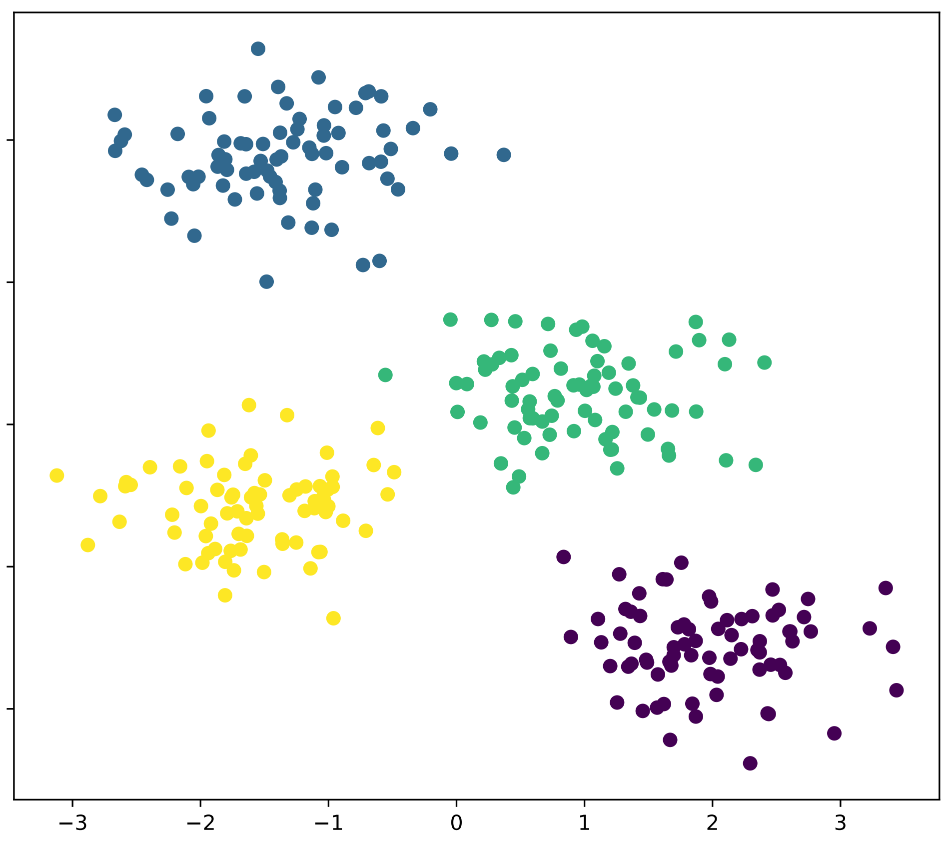
- Predict some continuous number from given input
- Linear Regression, NN, Gaussian Processes, …
- Housing prices, portfolio allocation, …

Classification



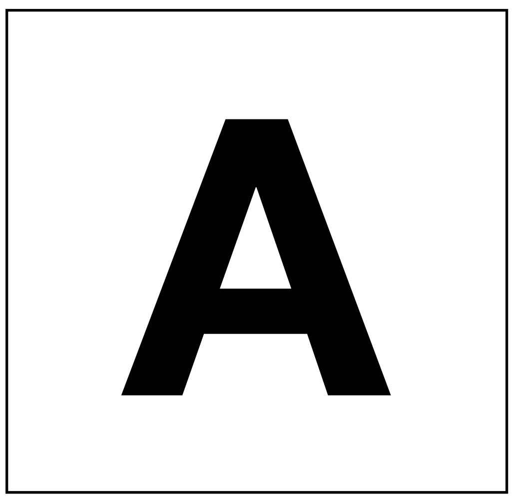
- Predict some discrete class labels from given input
- Logistic Regression, NN, SVM, Gaussian Process Classifier, …
- Image Recognition, financial crisis forecast, medical diagnosis …

Clustering

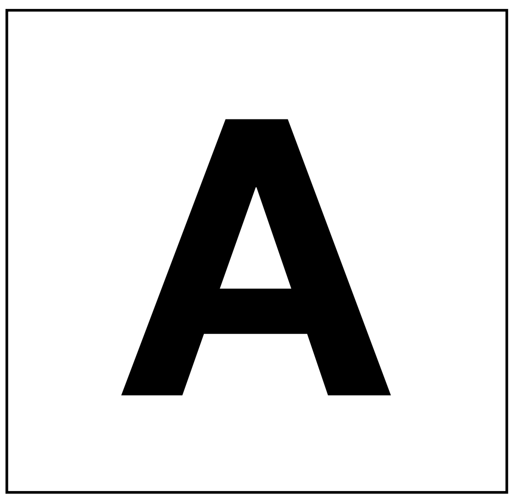


- Group together similar points from given input
- k-means, Gaussian mixture, ...
- Recommender systems, document modelling, …

Dimensionality Reduction

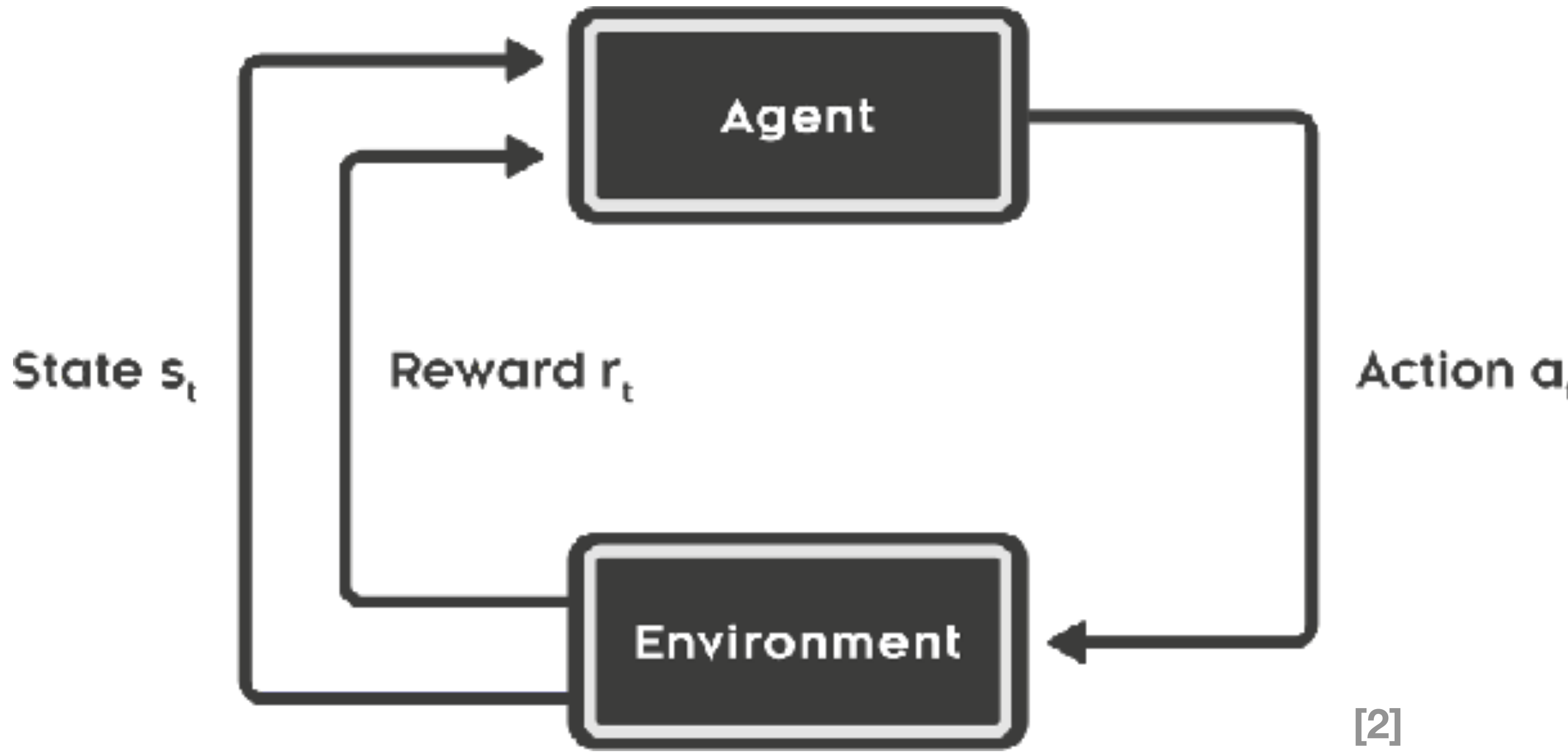


Dimensionality Reduction



- Reduction of data dimensionality while keeping information high
- PCA, FA, MDS, Isomap, GPLVM, ...
- Data preprocessing, compression, visualization …

Reinforcement Learning



[2]

- Sequential decision making to maximize future reward
- Q-learning, SARSA, ...
- Robotics, games, trading …

[2]: reinforce.io/reinforcement-learning

Neural Networks

- Linear Regression $y = Wx + b$
- Logistic Regression $y = \sigma(Wx + b)$
- How do we get some non-linearity in there?

Neural Networks

- Linear Regression $$y = Wx + b$$
- Logistic Regression $$y = \sigma(Wx + b)$$
- How do we get some non-linearity in there?
  - get more non-linear basis functions
  $$x \to (\Phi_1(x), \Phi_2(x), \ldots, \Phi_D(x)) = \Phi$$
  - build hierarchical models
  $$y = \sigma(Wx + b)$$
  $$\to W_2\sigma_1(W_1x + b_1) + b_2$$
  $$\to W_3\sigma_2(W_2\sigma_1(W_1x + b_1) + b_2) + b_3$$

Probabilistic Machine Learning

- We are living in a really simple world
- only known (data) and unknown (hypothesis) quantities exist

$$P(hypothesis \mid data) = \frac{P(data \mid hypothesis) P(hypothesis)}{P(data)}$$

Probabilistic Machine Learning

- We are living in a really simple world
- only known (data) and unknown (hypothesis) quantities exist

$$P(hypothesis \mid data) = \frac{P(data \mid hypothesis)P(hypothesis)}{P(data)}$$

- Inference (Learning)

$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)}$$

- Prediction

$$P(x \mid D) = \int P(x \mid \theta)P(\theta \mid D)\,d\theta$$

Probabilistic Machine Learning

- We are living in a really simple world
- only known (data) and unknown (hypothesis) quantities exist

$$P(hypothesis \mid data) = \frac{P(data \mid hypothesis)P(hypothesis)}{P(data)}$$

- Inference (Learning)
- Prediction

Probabilistic Programming

Probabilistic Machine Learning

- We are living in a really simple world

- only known (data) and unknown (hypothesis) quantities exist

$$P(hypothesis \mid data) = \frac{P(data \mid hypothesis) P(hypothesis)}{P(data)}$$

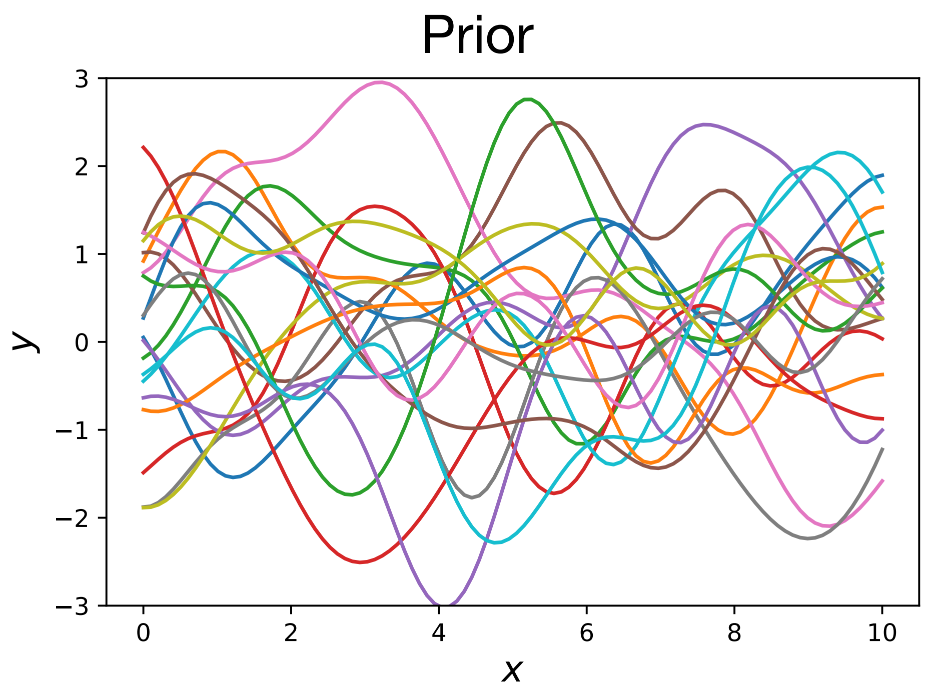
- Inference (Learning)

- Prediction

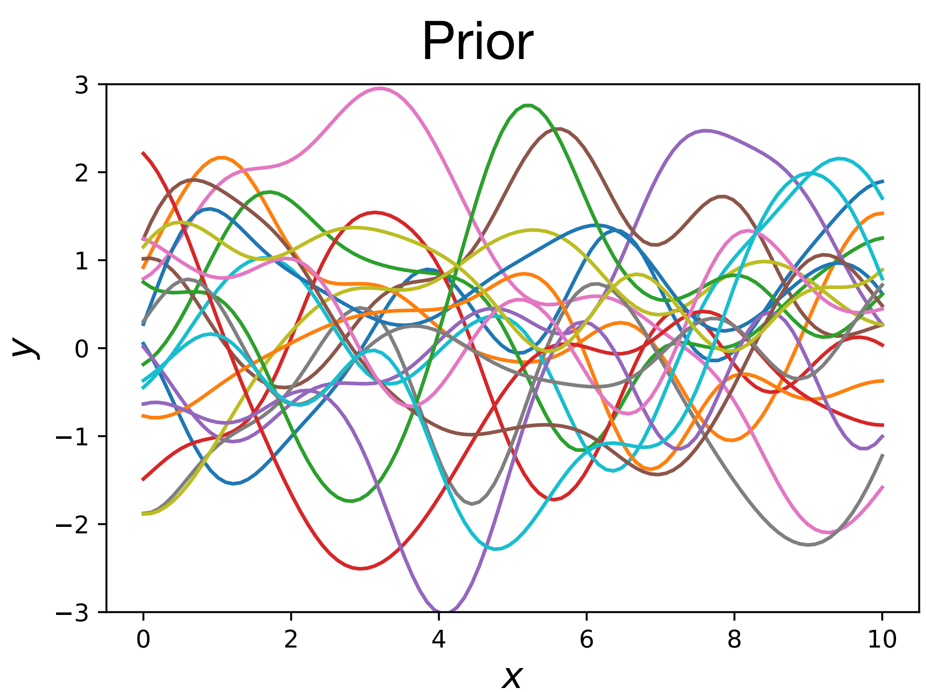
Probabilistic Programming

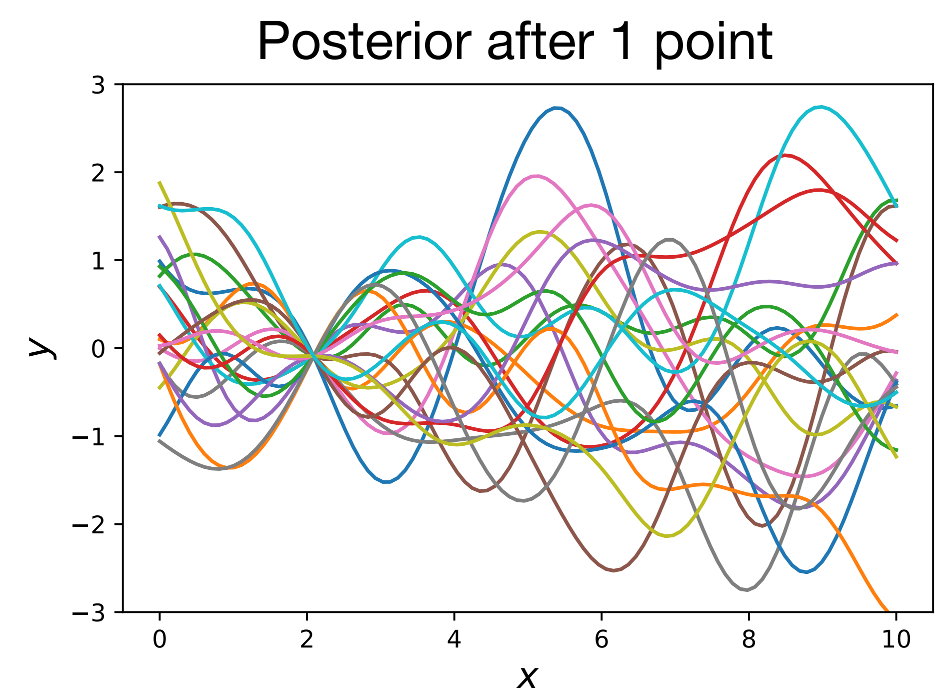
Edward, Stan, PyMC3, Pyro, ...

Probabilistic Machine Learning

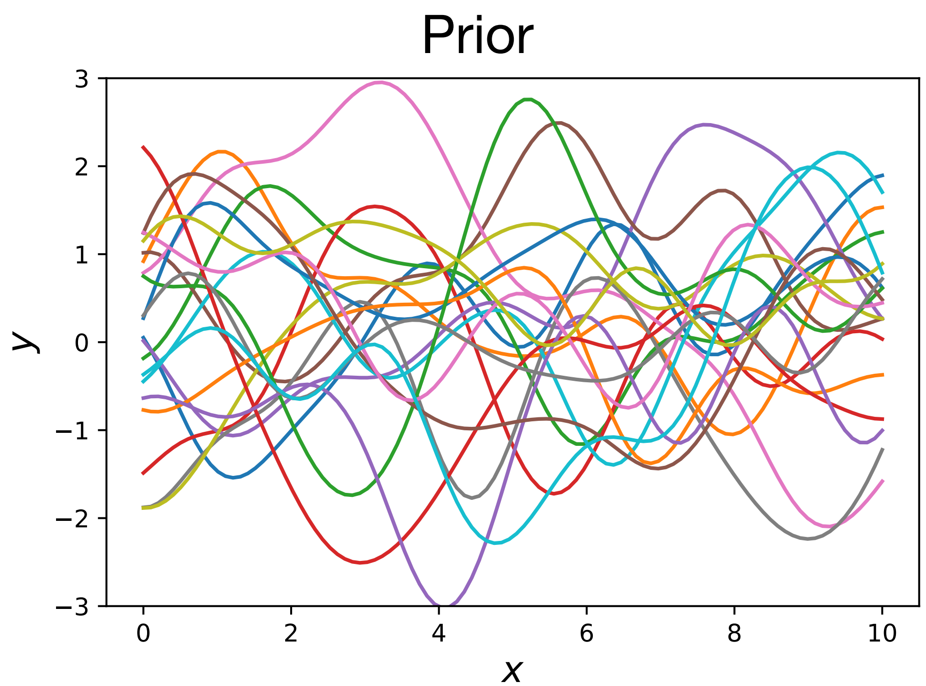


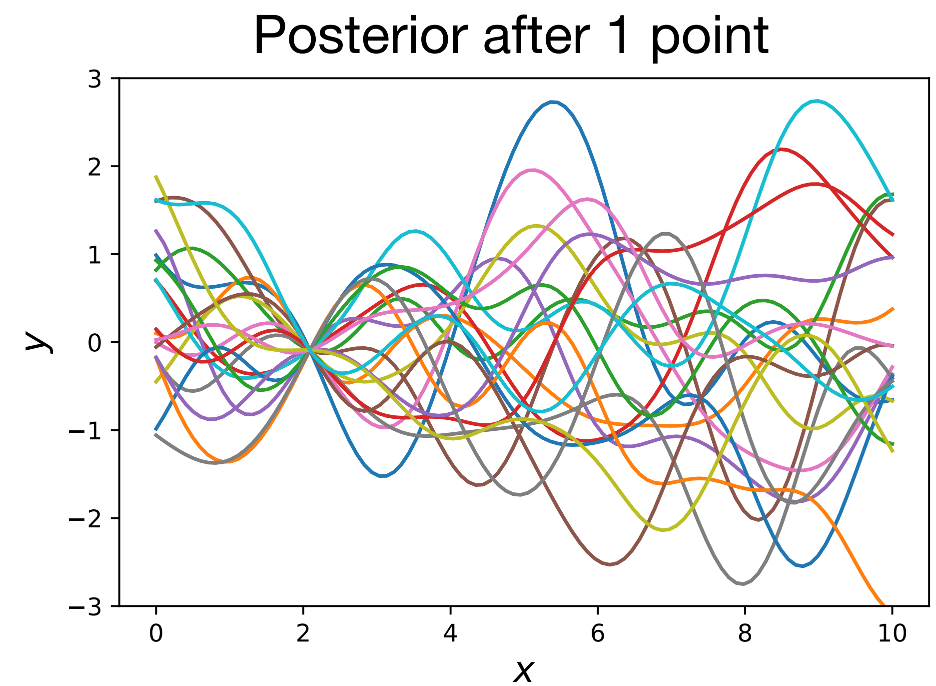
Probabilistic Machine Learning

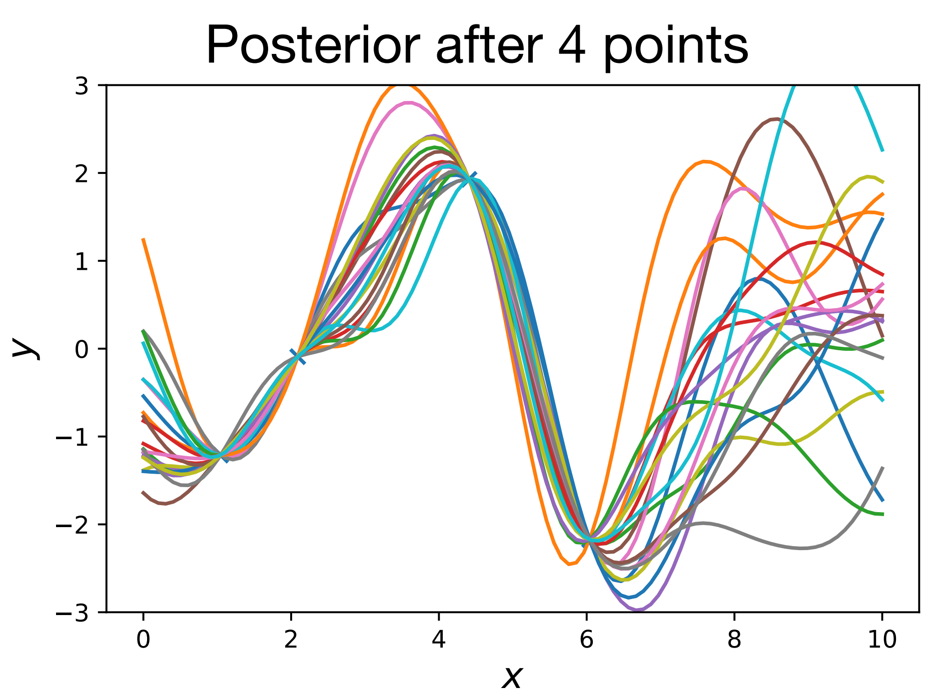




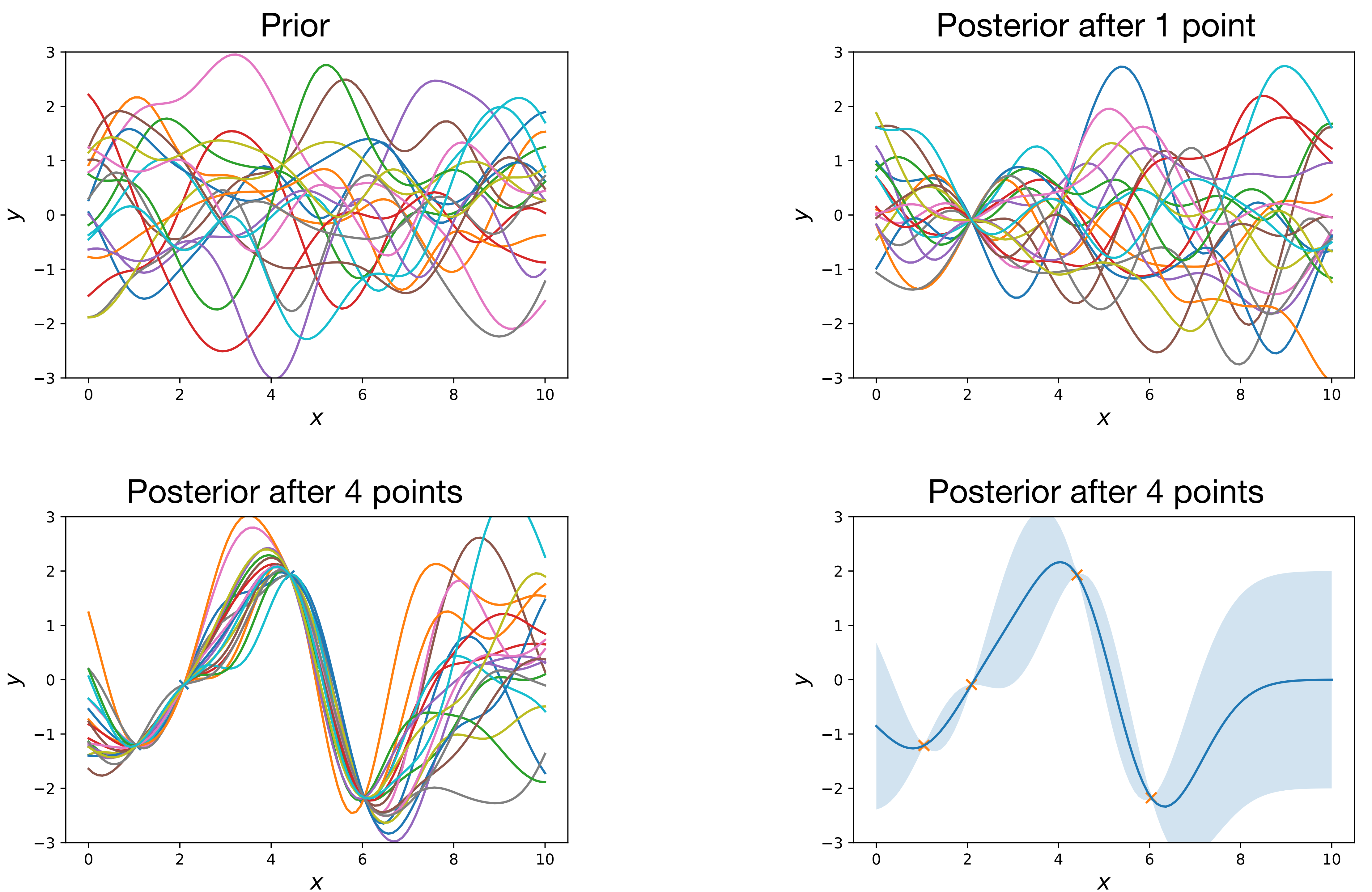
Probabilistic Machine Learning



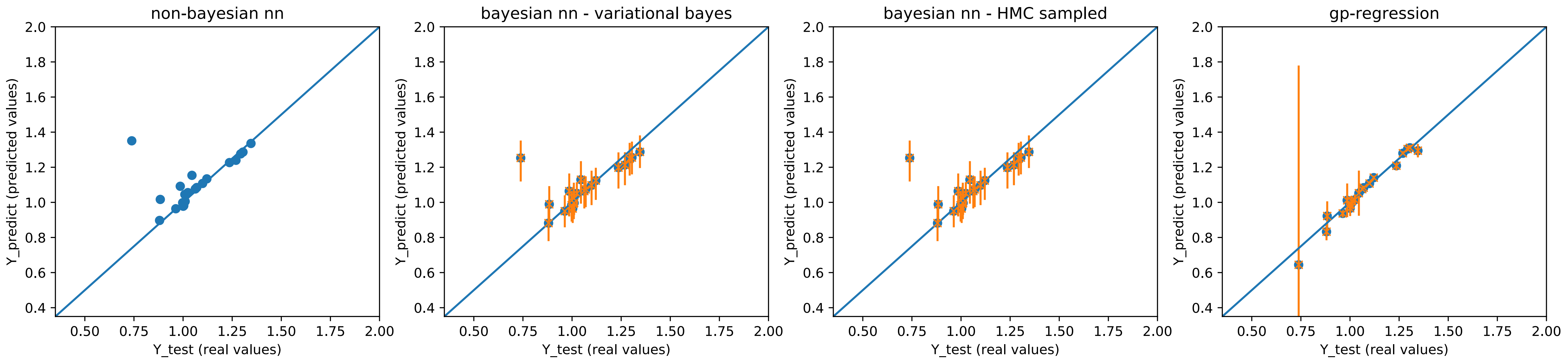




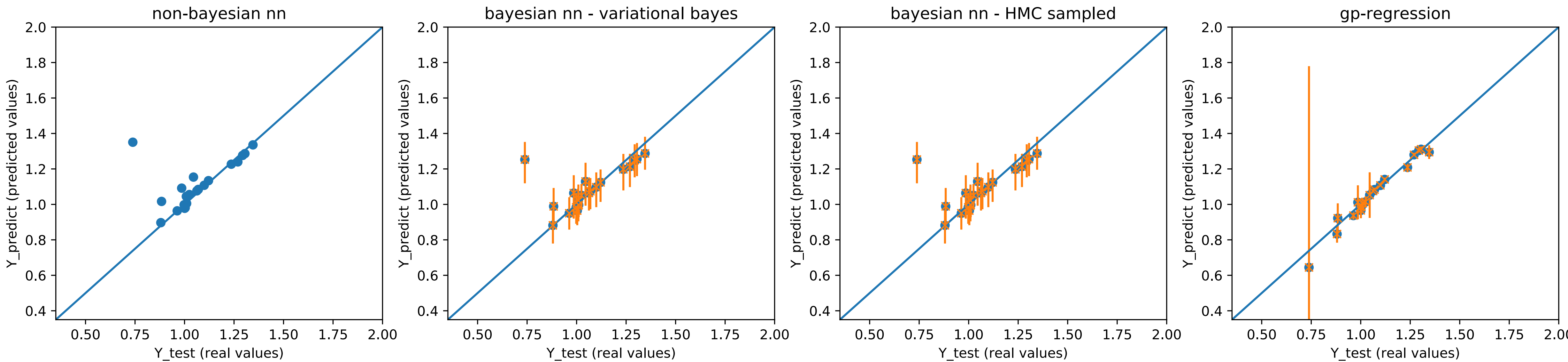
Probabilistic Machine Learning



Model Comparison



Model Comparison



Conclusion

- We need to build models that know when they don’t know!

Thanks for your attention!