Reinforcement Learning with Misspecified Bayesian Nonparametric Model Classes

Joshua Joseph, Alborz Geramifard, Jonathan P. How and Nicholas Roy

**Bayesian Nonparametric RBMS**

- Unlike the parametric approach, the predictions from a Bayesian nonparametric model is a function of the training data.
- To adapt parametric RBMS to Bayesian nonparametric model classes:
  - Policy evaluation using [2] still works
  - Policy improvement is unclear
  
For Bayesian nonparametric models, what does this term mean?

**Results on a Toy Problem**

Domain description:
- Actions = \{up, right\}
- Gaussian process dynamics model
- -1 reward for each time step
- -100 for falling in a pit
- Taking actions on the ice results in "slipping" south
- 100 episodes of training data from a random policy

The gradient for RBMS policy improvement was computed by randomly adding and removing data from the Gaussian process

**Conclusion**

- RBMS is able to learn models from misspecified model classes that perform well in cases where learning based on minimizing prediction error does not
- RBMS allows us to use smaller model classes, resulting in significantly lower sample complexity than using larger, more expressive model classes
- The extension of RBMS to Bayesian Nonparametric models is promising but requires more work to understand how to perform policy improvement
