# (Machine) Learning Robot Control Policies

**Daniel H Grollman, Odest Chadwicke Jenkins**
{dang,cjenkins}@cs.brown.edu

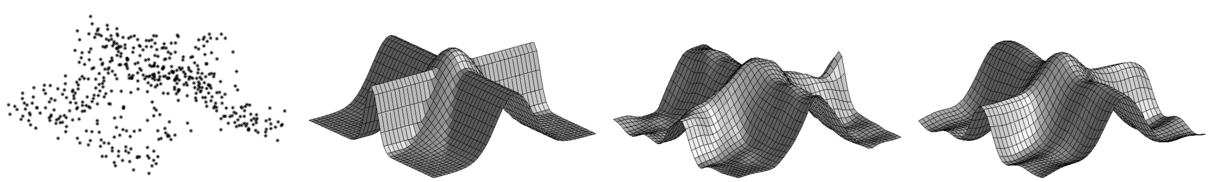
It currently requires years of education and practice before a skilled user can successfully program a sophisticated robot platform to perform a given task. We are exploring ways in which statistical machine learning techniques can enable **Learning from Demonstration**, an approach where users 'reprogram' a robot without writing code. In this scenario, a user demonstrates the desired task and the robot learns to perform the task by observing its performance. We treat this learning as a form of **Policy Transfer**, where the decision making policy latent in the demonstrator is transitioned onto the robot.

Taking perception and motion processes as fixed, we represent each policy as a functional mapping from perceived states to desired actions ($\pi(\hat{s}) \rightarrow a^*$). Using teleoperation, a demonstrator guides the robot through an instance of the desired behavior, creating a set of matched inputs and outputs. Function approximation techniques can then be applied to find an approximation of the control policy ($\hat{\pi}$).

We have left the tasks undefined, as we are interested in how robots can be made to learn **Unknown Tasks**, tasks not predefined during construction and original programming. Robots that exhibit **Lifelong Learning**, learning over extended periods (years) and in multiple domains, will likely need to deal with this issue. We have thus been exploring nonparametric function approximators. In addition, by using an algorithm capable of fast ($\sim$ 30Hz) inference and prediction on our system, we can enable interactive tutelage, where a demonstrator can observe and correct a learned behavior in realtime using **Mixed-Initiative Control**.

In our work so far [2], we have explored two such algorithms: Locally Weighted Projection Regression (LWPR) [4] and Sparse Online Gaussian Processes (SOGP) [1]. Figure 1 compares these two algorithms on a synthetic data set. Our initial robot-based experiments have focused on soccer-related tasks with robot dogs, and have shown successful learning from both hand-coded controllers and human demonstration.

Currently, we assume that the desired mapping ($\pi$) is functional, that each input has only one correct output. This is not the case in all contexts, as a robot may be able to perform two or more task-appropriate actions. We are interested in techniques that can learn such non-deterministic mappings directly from input-output pairs, such as mixtures of experts [3]. In addition, by incorporating aspects of reinforcement learning we hope to further our ability to perform task performance refinement and task structure learning.



(a) Noisy Data (b) Ground Truth (c) LWPR, MSE = 0.0202 (d) SOGP, MSE = 0.0244

Figure 1: SOGP and LWPR with default parameters compared on the cross function. We limit SOGP's capacity to the number of receptive fields used by LWPR (22).

## References

[1] Lehel Csató and Manfred Opper. Sparse online gaussian processes. *Neural Computation*, 14(3):641–669, 2002.

[2] Daniel H Grollman and Odest Chadwicke Jenkins. Learning robot soccer skills from demonstration. In *6th IEEE Intl. Conference on Development and Learning*, 2007.

[3] Edward Meeds and Simon Osindero. An alternative infinite mixture of gaussian process experts. In *Neural Information Processing Systems (NIPS)*, pages 883–890, 2006.

[4] Sethu Vijayakumar, Aaron D'Souza, and Stefan Schaal. Incremental online learning in high dimensions. *Neural Computation*, 17(12):2602–2634, 2005.