Using Deep Belief Nets to Learn Covariance Kernels for Gaussian Processes
R. Salakhutdinov and G. Hinton, University of Toronto, Poster IDM41

- Kernel-based methods usually specify a fixed type of kernel in advance and only adapt a few hyper-parameters. They do not learn a complicated task-specific kernel. This is wasteful if there is a lot of unlabeled data and only a little labeled data: On one of the regression tasks we use to compare methods it gives: fixed kernel: 16.3% error

- We learn a deep belief net (DBN) on a big unlabeled dataset and then use the features in the deepest layer to train a Gaussian Process (GP) on the labeled data. This reduces the error: greedily learned kernel: 11.2% error

- Then we back-propagate derivatives from the GP to adapt the features in every layer of the DBN. This produces much better performance: fine-tuned kernel: 6.4% error