

CRUNCH Seminars at Brown, Division of Applied Mathematics

Friday – August 28, 2018

**Paper Review: Kernel Flows: from learning kernels from data into the abyss
(Houman Owhadi and Gene Ryan Yooy, [arXiv:1808.04475](https://arxiv.org/abs/1808.04475))**

Guofei Pang

Learning can be seen as approximating an unknown function by interpolating the training data. Kriging offers a solution to this problem based on the prior specification of a kernel. We explore a numerical approximation approach to kernel selection/construction based on the simple premise that a kernel must be good if the number of interpolation points can be halved without significant loss in accuracy (measured using the intrinsic RKHS norm $\|\cdot\|$ associated with the kernel). We first test and motivate this idea on a simple problem of recovering the Green's function of an elliptic PDE (with inhomogeneous coefficients) from the sparse observation of one of its solutions. Next we consider the problem of learning non-parametric families of deep kernels of the form $K_1(F_n(x); F_n(x'))$ with $F_{n+1} = (\text{Id} + \epsilon G_{n+1}) \circ F_n$ and $G_{n+1} \in \text{span}\{K_1(F_n(x_i); \cdot)\}$. With the proposed approach constructing the kernel becomes equivalent to integrating a stochastic data driven dynamical system, which allows for the training of very deep (bottomless) networks and the exploration of their properties. These networks learn by constructing flow maps in the kernel and input spaces via incremental data-dependent deformations/perturbations (appearing as the cooperative counterpart of adversarial examples) and, at profound depths, they (1) can achieve accurate classification from only one data point per class (2) appear to learn archetypes of each class (3) expand distances between points that are in different classes and contract distances between points in the same class. For kernels parameterized by the weights of Convolutional Neural Network, minimizing approximation errors incurred by halving random subsets of interpolation points, appears to outperform training (the same CNN architecture) with relative entropy and dropout.