July 2012

MADALGO seminar by Jeff M. Phillips, University of Utah

Discrepancy for Kernel Range Spaces

Abstract:

We study the worst case error of kernel density estimates via subset approximation. A kernel density estimate of a distribution is the convolution of that distribution with a fixed kernel (e.g. Gaussian kernel). Given a subset (i.e. a point set) of the input distribution, we can compare the kernel density estimates of the input distribution with that of the subset and bound the worst case error. If the maximum error is eps, then this subset can be thought of as an eps-sample (aka an eps-approximation) of the range space defined with the input distribution as the ground set and the fixed kernel representing the family of ranges. Interestingly, in this case the ranges are not binary, but have a continuous range (for simplicity we focus on kernels with range of [0,1]); these allow for smoother notions of range spaces.

It turns out, the use of this smoother family of range spaces has an added benefit of greatly decreasing the size required for eps-samples. For instance, in the plane the size is O((1/eps^{4/3}) log^{2/3}(1/eps)) for disks (based on VC-dimension arguments) but is only O((1/eps) sqrt{log (1/eps)}) for Gaussian kernels and for kernels with bounded slope that only affect a bounded domain. These bounds are accomplished by studying the discrepancy of these "kernel" range spaces, and here the improvements in bounds are even more pronounced. In the plane, we show the discrepancy is O(sqrt{log n}) for these kernels, whereas for balls there is a lower bound of Omega(n^{1/4}).