Kernels as Features: On Kernels, Margins, and Low-dimensional Mappings

Avrim Blum

Carnegie Mellon University, Pittsburgh PA 15213-3891, USA.

Kernel functions are typically viewed as providing an implicit mapping of points into a high-dimensional space, with the ability to gain much of the power of that space without incurring a high cost if data is separable there by a large margin. However, the Johnson-Lindenstrauss lemma suggests that in the presence of a large margin, a kernel function can also be viewed as a mapping to a *low*-dimensional space if we can generate random vectors in the implicit space to perform the random projection. In this talk, I will discuss how we can efficiently produce such low-dimensional mappings, given only black-box access to the kernel function. That is, given just a program that computes $K(x, y)$ on inputs x,y of our choosing, can we construct an explicit (small) set of features that effectively capture the power of the implicit high-dimensional space? Our method requires black-box access to the underlying data distribution (i.e., unlabeled examples) and can be viewed as a way of converting any given kernel into a distribution-dependent feature set; we also show that without access to the data distribution, then this is not possible in general for an arbitrary black-box kernel.

This is joint work with Nina Balcan and Santosh Vempala.