



THE WEIZMANN INSTITUTE OF SCIENCE  
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Machine Learning and Statistics Seminar

Room 1 ,Ziskind Building  
on Wednesday, Feb 08, 2017  
at 11:15

Haim Avron  
Tel Aviv University

Large-scale and Non-approximate Kernel Methods Using Random Features

Abstract:

Kernel methods constitute a mathematically elegant framework for general-purpose infinite-dimensional non-parametric statistical inference. By providing a principled framework to extend classical linear statistical techniques to non-parametric modeling, their applications span the entire spectrum of statistical learning. However, training procedures naturally derived via this framework scale poorly and with limited opportunities for parallelization. This poor scalability poses a significant barrier for the use of kernel methods in big data applications. As such, with the growth in data across a multitude of applications, scaling up kernel methods has acquired renewed and somewhat urgent significance. Random feature maps, such as random Fourier features, have recently emerged as a powerful technique for speeding up and scaling the training of kernel-based methods. However, random feature maps only provide crude approximations to the kernel function, so delivering state-of-the-art results requires huge amount of random features. Nevertheless, in some cases, even when the number of random features is driven to be as large as the training size, full recovery of the generalization performance of the exact kernel method is not attained. In the talk I will show how random feature maps can be used to efficiently perform non-approximate kernel ridge regression, and thus there is no need to compromise between quality and running time. The core idea is to use random feature maps to form preconditioners to be used in solving kernel ridge regression to high accuracy. I will describe theoretical conditions on when this yields an effective preconditioner, and empirically evaluate the method and show it is highly effective for datasets of up to one million training examples.