Random Features for Large-Scale Kernel Machines

The paper *"Random Features for Large-Scale Kernel Machines"* by Ali Rahimi and Ben Recht introduces a novel approach to accelerate the training of kernel machines. Traditional kernel machine training, exemplified by Support Vector Machines (SVMs), suffers from scalability issues due to prolonged training times with increasing dataset sizes. The authors propose a method that maps input data to a randomized low-dimensional feature space and applies fast linear methods. This innovative mapping approximates the inner products in the transformed data space to those in the kernel space defined by a user-specified shift-invariant kernel. This strategy leverages linear models' computational efficiency while retaining kernel methods' expressive power.

The methodology centers on two types of randomized feature maps: Random Fourier Features and Random Binning Features. These maps are designed to approximate the inner product between data points in a transformed space, aiming to mimic the computation in the kernel space.

Random Fourier Features are derived by sampling sinusoids from the Fourier transform of the kernel function, effective for interpolation tasks. The key mathematical insight behind this approach is that the inner product of the transformed data points serves as an unbiased estimator of the kernel. The mapping is defined as:

$$z_\omega(x) = \sqrt{2}\cos(\omega x + b), \tag{1}$$

where ω is drawn from the Fourier transform $p(\omega)$ of the kernel $k(\delta)$, and b is uniformly drawn from $[0, 2\pi]$. This approach relies on the property that the expected value of the product of two transformed points approximates the desired shift-invariant kernel.

The Random Binning Features method partitions the input space using randomly shifted grids at various resolutions. This technique is particularly suited for kernels that depend on the L_1 distance between data points. It preserves locality by assigning binary bit strings to points based on their bins, with the inner product of these strings estimating the kernel evaluation. The fundamental idea is that the probability of two points being assigned to the same bin is proportional to the kernel evaluation of those points.

Experimental results demonstrate the effectiveness of these randomized feature maps in reducing the computational burden of training while maintaining, or even improving, the performance of kernel machines on large-scale classifi-

cation and regression tasks. The experiments showcase that linear algorithms equipped with these features can compete favorably against state-of-the-art kernel machines, offering a significant reduction in training and evaluation time. This finding opens up new avenues for employing kernel-based methods in scenarios where computational resources are limited or when dealing with very large datasets.

In conclusion, Rahimi and Recht's work on randomized features for kernel machines marks a significant advancement in the field of machine learning, offering a practical solution to the scalability challenges of kernel-based methods. The proposed feature maps not only facilitate faster training and evaluation but also extend the applicability of kernel machines to larger datasets, potentially benefiting a wide range of applications in data science and beyond. The paper also suggests future research directions, including the exploration of combining different feature types and applying these methods to other learning tasks, indicating a promising area for further developments.