Abstract

Kernel Support Vector Machines deliver the state-of-the-art classification, but do not scale well for large datasets since the training and testing procedures operate directly on the $N \times N$ Gram matrix, where N is the number of training examples. In contrast, linear Support Vector Machines can train and test samples fast even when data size is big on limited resources(e.g. personal PCs). Inspired by this, two algorithms, namely random Fourier features and the Nyström method, offer alternatives that speed up kernel methods by transforming the non-linear kernel problem to a linear one. We aim to study the theoretical and empirical differences between them. We report experiments on the MNIST dataset and show that both algorithms compare favourably to kernel methods in terms of accuracy and efficiency.