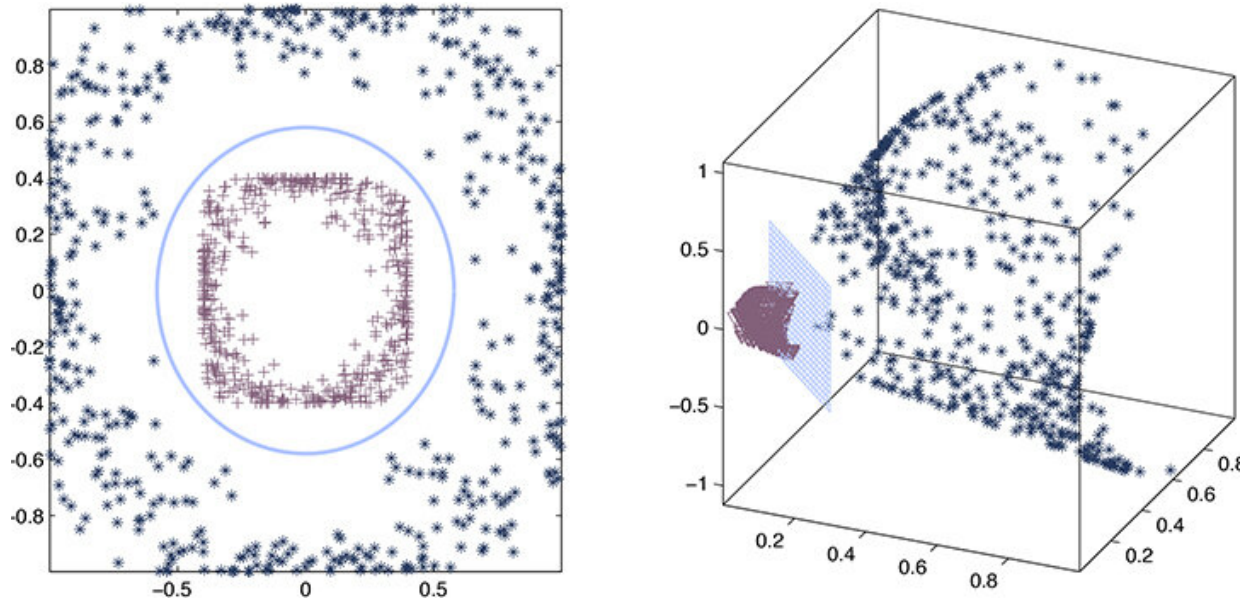# Approximating a kernel of truth

March 10 2020

The researchers used a process of error estimation and mathematical approximation to prove that their approximate kernel remains consistent with the accurate kernel. Credit: 2020 Ding et al.

By using an approximate rather than explicit "kernel" function to extract relationships in very large data sets, KAUST researchers have been able to dramatically accelerate the speed of machine learning. The approach promises to greatly improve the speed of artificial intelligence (AI) in the era of big data.

When AI is exposed to a large unknown data set, it needs to analyze the

data and develop a model or function that describes the relationships in the set. The calculation of this function, or kernel, is a computationally intensive task that increases in complexity cubically (to the power of three) with the size of the data set. In the era of big data and increasing reliance on AI for analysis, this presents a real problem where kernel selection can become impractically time consuming.

With the supervision of Xin Gao, Lizhong Ding and his colleagues have been working on methods to speed up kernel selection using statistics.

"The computational complexity of accurate kernel selection is usually cubic with the number of samples," says Ding. "This kind of cubic scaling is prohibitive for big data. We have instead proposed an approximation approach for kernel selection, which significantly improves the efficiency of kernel selection without sacrificing predictive performance."

The true or accurate kernel provides a verbatim description of relationships in the data set. What the researchers found is that statistics can be used to derive an approximate kernel that is almost as good as the accurate version, but can be computed many times faster, scaling linearly, rather than cubically, with the size of the data set.

To develop the approach, the team had to construct specifically designed kernel matrices, or mathematical arrays, that could be computed quickly. They also had to establish the rules and theoretical bounds for [selection](#) of the approximate kernel that would still guarantee learning performance.

"The main challenge was that we needed to design new algorithms satisfying these two points at the same time," says Ding.

Combining a process of error estimation and mathematical

approximation, the researchers were able to prove that their approximate kernel remains consistent with the accurate kernel and then demonstrated its performance in real examples.

"We have shown that approximate methods, such as our computing framework, provide sufficient accuracy for solving a kernel-based learning method, without the impractical computational burden of accurate methods," says Ding. "This provides an effective and efficient solution for problems in data mining and bioinformatics that require scalability."

Provided by King Abdullah University of Science and Technology