Despite the popularity and superior performance of Gaussian-kernel support vector machine (SVM), the two hyperparameters sigma (scale) and C (tradeoff) remain hard to be tuned. Many techniques have been developed to address this challenge such as grid search (full/random) and Bayesian methods. However, they all ignore the intrinsic geometry of training data as well as the interpretation of the two parameters and tackle the problem solely from an optimization point of view, thus being computationally expensive. A few methods such as the Jaakkola and Caputo heuristics try to learn the sigma parameter directly from training data. Unfortunately, both of them require computing distances between points from different training classes and also sorting them which can be a great computational burden for large data sets.

In this poster we present two novel techniques for efficiently selecting sigma and C, respectively: (1) For the sigma parameter we point out that it corresponds to the local scale of the training classes and propose a fast nearest-neighbor approach for directly setting its value. (2) For the C parameter, we have observed in many cases that the validation accuracy peaks at some C or stabilizes after some C (as C is set to increase). This motivated us to propose a procedure that starts from the lower end of the C grid and detects an elbow point on the validation accuracy curve (which is gradually computed). Such a choice of C leads to robust margins and avoids testing the bigger C values which tend to be computationally more expensive.

We have tested our combined algorithm against the existing approaches on a number of data sets and obtained very favorable results. In most cases the predictive accuracy of our method is at least comparable to the best of those methods, though our method uses only a single value of sigma. In addition, our method is much faster than the other methods. Finally, our method is insensitive to the number k of the nearest neighbors used, and in general any k between 6 and 10 gives competitive results.