ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Kernel Thinning for faster KSVM hyper-parametrization

Blanca Cano, Ángela Fernández, José R. Dorronsoro

Departamento de Ingeniería Informática – Universidad Autónoma de Madrid, Spain

Abstract. This work presents *KT-Funnel*, a novel procedure to accelerate hyperparameter tuning in kernel methods, together with an empirical study of efficient cross-validation strategies for Kernel Support Vector Machines (KSVM) based on Kernel Thinning (KT). The main aim is to reduce computational cost while preserving predictive accuracy. Experiments on 25 classification datasets show that *KT* and *KT-Funnel* significantly speed up hyperparameter tuning, being at least twice faster than traditional cross-validation. In particular, the proposed method with higher thinning levels attains comparable balanced accuracy and improved hyperparameter ranking stability, demonstrating its scalability and reliability for KSVM model selection.

1 Introduction

Kernel Support Vector Machines (KSVM) [1] remain highly effective for many learning tasks, but their computational cost—at least quadratic in the dataset size—becomes prohibitive for large-scale problems [2], particularly during cross-validation (CV) [1], where multiple training runs are required.

To overcome this limitation, we propose a new procedure called *Kernel Thinning Funnel (KT-Funnel)* to accelerate hyperparameter tuning for kernel machines. This method builds upon Kernel Thinning (KT) [3, 4, 5], which reduces the training sample size while preserving the data distribution, thereby lowering the cost of each CV evaluation.

We experimentally study both KT and the new introduced approach, which performs a progressive, multi-stage reduction of the dataset and search space across consecutive thinning levels. Our goal is to determine whether these strategies can find near-optimal hyperparameters, how much the sample can be reduced without accuracy loss, and the computational gains achieved.

Results across multiple datasets show that strategies based on KT—especially the "funnel" approach—preserve balanced accuracy while reducing computation time. The paper is organized as follows: Section 2 briefly reviewed KT, Section 3 introduces the *KT-Funnel* method, Section 4 describes the experimental results, and Section 5 summarizes out findings and directions for future work.

2 Kernel Thinning

KT is a recent technique designed to compress a distribution P more efficiently than classical i.i.d. sampling or standard thinning [3]. Given an admissible kernel k, the method reduces a set of n samples from P into a smaller subset of size

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

$\lfloor \frac{n}{2^m} \rfloor$ (with m a fixed integer), while preserving a comparable Maximum Mean Discrepancy (MMD) [6] in the associated Reproducing Kernel Hilbert Space [7]. The overall complexity of the procedure is $\mathcal{O}(n^2)$. Beyond the base kernel k, KT employs an auxiliary function referred to as the *square-root kernel*, defined as follows:

Definition 2.1 (Square-root Kernel) *Let* $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ *be a symmetric, positive-definite kernel. A function* $k_{\mathrm{rt}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ *is called a* square-root kernel *for* k *if* $k_{\mathrm{rt}}(x, \cdot)$ *is square-integrable for all* $x \in \mathbb{R}^d$*, and*

$$k(x, y) = \int_{\mathbb{R}^d} k_{\mathrm{rt}}(x, z) k_{\mathrm{rt}}(y, z)\, dz \quad \textit{for all } x, y \in \mathbb{R}^d.$$

For example, when k is the Gaussian kernel $k(x, y) = \exp\left(-\gamma \|x - y\|^2\right)$, a valid square-root kernel is $k_{rt}(x, y) = \left(\frac{4\gamma}{\pi}\right)^{d/4} \exp(-2\gamma \|x - y\|^2)$.

The KT procedure consists of two main components [3]. The first, KT-Split, recursively partitions the dataset into 2^m candidate subsets of size roughly $n/2^m$ using randomized halving based on the square-root kernel k_{rt}, ensuring balanced and diverse splits. The second, KT-Swap, identifies the candidate subset that minimizes MMD relative to the full dataset, and refines it through a greedy swapping procedure with respect to the target kernel k. Together, these steps produce a reduced coreset of $\lfloor \frac{n}{2^m} \rfloor$ elements with strong theoretical guarantees in terms of MMD; see [3] for more details.

3 Surrogate hyperparametrization via Kernel Thinning

KT can be effectively integrated into a surrogate hyperparameter tuning procedure for kernel methods [8] to substantially reduce its computational cost. The method searches for the optimal hyperparameters by evaluating all possible hyperparameter combinations using a subset of size $n/2^m$ from each dataset, where the subset is selected through the *KT* procedure. The parameter m controls the thinning level, determining the size of the coreset used for validation. We refer to this approach as *KT*.

A drawback of this method is that performance deteriorates as m increases. To mitigate this effect, we propose the **Kernel Thinning Funnel** (KT-Funnel) procedure, described in Algorithm 1. In this approach, the set of candidate hyperparameters is iteratively refined through m_{max} thinning stages. At each stage t ($1 \leq t \leq m_{max}$), a subset of size $n/2^t$ of the training data is selected using the KT procedure, and the model is evaluated to discard the worst-performing hyperparameter combinations. After completing all m_{max} stages, the remaining reduced set of promising hyperparameters is finally re-evaluated using the full dataset to identify the optimal configuration.

KT-Funnel improves computational efficiency compared to standard CV by progressively balancing the reduction of poor hyperparameters with the increasing reliability of performance estimates as the subset size grows. Formally, let

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Algorithm 1 Kernel Thinning Funnel with Mixed Hyperparameters

Require:
- $m_{max} = m_t > \cdots > m_1 > m_0 = 0$,
- retention ratio $\alpha \in]0, 1[$,
- initial candidates set Λ_0,
- **hyperparameter decomposition** $\Lambda = \Lambda^K \times \Lambda^{\neg K}$, where Λ^K are the hyperparameters that only affects the kernel an it size is n_k, while $\Lambda^{\neg K}$ the ones that does not affect the kernel and its size is n_c.
- Kernel functions $k(\cdot,\cdot)$ and $k_{\mathrm{rt}}(\cdot,\cdot)$ parameterized by λ^K,
- input points X of size n,
- a validation function $\mathit{validation} : \Lambda \times X \to \mathbb{R}_0^+$,

```
Initialize an empty cache C                          ▷ Store KT subsets by key (λ^K, m)
for i ∈ {0, ..., t − 1} do
    R ← ∅
    for each λ = (λ^K, λ^¬K) ∈ Λ_i do
        if (λ^K, m_{t−i}) ∉ C then
                         ▷ Compute the full hierarchy {S_{m_j}^{(λ^K)}}_{j=1}^t in one pass
            C[λ^K] ← KT_split(X, k)
        end if
        S ← C[(λ^K, m_{t−i})]            ▷ Get stored KT subset of X size of n/2^m_{t−i}
        score ← validation(λ, S)
        R ← R ∪ {(score, λ)}
    end for
    Λ_{i+1} ← { λ : (s, λ) ∈ R and s is among the top α results }
end for
λ̂ ← arg max_{λ∈Λ_t} validation(λ, X)                   ▷ Final evaluation on full X
return Best estimated hyperparameter λ̂
```

n_k and n_c denote the number of kernel-related and kernel-independent hyperparameter values, respectively.

The standard KSVM CV cost scales as $\Omega\left(n_k n_c n^2\right)$, while for KT-Funnel strategy is $\Omega\left(n^2 n_k + n_c n_k n^2 \sum_{i=0}^{t} \alpha^i 2^{-2m_{t-i}}\right)$, where the thinning levels are $m_{max} = m_t > \cdots > m_1 > m_0 = 0$ and $\alpha \in]0, 1[$ represents the fraction of top configurations retained per stage. In practice, with consecutive thinning levels for $\alpha \leq 0.75$ and $m_{max} \geq 2$, computational cost improvements are observed when $n_c \geq 2$, indicating that the KT-Funnel procedure yields substantial efficiency gains in realistic scenarios. For our experiments we fixed $\alpha = 0.5$ and maximum deeper $m_{max} = 4$ which will provide a theoretical improvement of at least eight times. We will refer to this procedure as *KT-Funnel*, or simply as *Funnel*.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Table 1: Position of the best hyperparameter configurations in the global ranking derived from the full KSVM evaluation (lower is better).

	KT				Funnel			
m_{max}	1	2	3	4	1	2	3	4
mean	17	22	26	34	3	6	9	12
std	15	18	21	23	5	9	11	14
median	15	15	17	34	1	1	5	6
min	1	1	1	2	1	1	1	1
max	52	61	62	77	18	30	38	45

4 Experiments

In this section, we present the experiments designed to compare the computational cost and predictive performance of Gaussian KSVM using standard CV procedure and the KT-based methods, KT and $Funnel$, introduced in Section 3. Experiments were conducted on 25 binary and multiclass classification datasets from the UCI repository (IDs 17, 19, 27, 53, 59, 73, 76, 80, 94, 101, 109, 110, 144, 159, 267, 327, 468, 519, 529, 544, 545, 563, 603, 850, 863), covering a broad range of sizes and distributions, as also studied in [8]. Experiments were run on a shared server with Intel® Xeon® E5-2680 v4 CPUs and 755 GiB RAM.

The reported training time and balanced accuracy in test correspond to the mean values obtained from a stratified nested CV with five outer and five inner folds. Balanced accuracy is used to prevent bias caused by class imbalance and sampling variability across the diverse datasets considered. The search space for the kernel hyperparameter γ includes 10 values spaced logarithmically between $2^{-2}/d$ and $2^{7}/d$, where d is the dimensionality of the dataset, while the regularization parameter C takes 10 values spaced logarithmically between 10^{-2} and 10^{5}. The combination of both hyperparameters yields 100 CV configurations.

Within the inner loops of the CV, mean balanced accuracy value for validation is obtained for each hyperparameter candidate. These scores determine the relevance ranking of the hyperparameters. Therefore, we can consider the ranking positions achieved by the best candidates from the KT and Funnel procedures in the full-dataset CV.

Table 1 shows the ranking positions that the best hyperparameter configurations of each method would occupy within KSVM ranking over all the samples, evaluated over the 100 combinations. For all values of m, the Funnel procedure achieves lower (better) mean and median ranks than the corresponding KT configurations and exhibits smaller standard deviations, indicating more consistent performance. At higher depths (e.g., $m_{max} = 4$), Funnel compensates for the degradation observed in KT as m increases.

The ranking trend is also reflected in the balanced accuracy reported in Table 2. Regardless of the value of m, the Funnel procedure achieves in most cases slightly higher mean and median balanced accuracies than the correspond-

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Table 2: Balanced accuracy in test for the best hyperparameters configuration obtained for the different methods.

	KSVM	KT				Funnel			
m_{max}		1	2	3	4	1	2	3	4
mean	90.5	89.9	89.5	89.0	88.8	90.0	89.8	89.5	89.8
std	9.8	10.7	11.4	11.0	10.9	10.9	11.2	12.4	12.0
median	93.3	93.7	94.5	92.9	92.6	94.7	94.7	94.5	94.4

Table 3: Training time (in ms) of all candidate estimators for each procedure.

	KSVM	KT				Funnel			
m_{max}		1	2	3	4	1	2	3	4
mean	1705	314	82	24	11	647	288	132	73
std	5215	917	224	67	34	1527	633	287	179
median	77	15	4	2	1	25	15	10	6
max	26465	4653	1110	336	174	6855	2558	1227	853

ing KT configurations. This observation is supported by a Wilcoxon test with Bonferroni correction at a significance level of 0.05, conducted to compare the prediction outputs on the test sets. The results indicate that, in general, the differences between KT and Funnel are statistically significant, whereas no significant differences are found between Funnel and KSVM in most datasets.

Finally, regarding training time for all candidate estimators, Table 3 shows in ms that all proposed methods are, on average, at least twice as fast as the traditional CV procedure. The fastest configurations are KT $m = 4$, KT $m = 3$, and Funnel $m_{max} = 4$, in that order. This result highlights Funnel $m_{max} = 4$ as the most promising alternative, since it surpasses all KT variants in ranking performance while maintaining a highly competitive computational cost. Additionally, the Funnel $m_{max} = 1$ achieves the best overall ranking and balanced accuracy among all methods, while still remaining faster than the KSVM baseline.

5 Conclusion and Future Work

In this work, we present an experimental study of KT and a new proposed method, KT-Funnel, evaluating their performance as surrogate techniques for KSVM CV. The experiments, conducted on 25 diverse datasets from the UCI repository, demonstrate that both KT and Funnel approaches substantially reduce computational cost while maintaining competitive predictive performance.

The results show that the KT-Funnel method consistently achieves better ranking positions for the selected hyperparameters than KT, exhibiting lower variability across datasets and folds. In terms of balanced accuracy, funnel methods achieve results close to those obtained with the full dataset, even for higher thinning levels. Moreover, training time analysis reveals that both KT and Funnel procedures offer significant speed ups compared to standard CV without notable losses in prediction quality. Among all tested configurations, KT-Funnel with $m_{max} = 4$ achieves the best balance between speed and reliability, whereas $m_{max} = 1$ provides the highest balanced accuracy and the most accurate hyperparameter approximation with 1/3 of the cost of full KSVM.

Future work will focus on extending these approaches to other kernel methods and regression tasks, and developing adaptive strategies to automatically select the optimal thinning level m.

The authors acknowledge financial support from project PID2022-139856NB-I00 funded by MCIN/AEI/10.13039/ 501100011033/ FEDER, UE and project IDEA-CM (TEC-2024/COM-89) from the CAM, the Cátedra UAM-IIC de Ciencia de Datos y Aprendizaje Automático, and FPI-UAM, as well as computational support from the Centro de Computación Científica Universidad Autónoma de Madrid (CCC-UAM).

References

[1] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd edition edition, 2009.

[2] Meng Wang, Weijie Fu, Xiangnan He, Shijie Hao, and Xindong Wu. A survey on large-scale machine learning. *IEEE Transactions on Knowledge and Data Engineering*, 34(6), 2020.

[3] Raaz Dwivedi and Lester Mackey. Kernel Thinning. *JMLR*, 25(152):1–77, 2024.

[4] Raaz Dwivedi and Lester Mackey. Generalized Kernel Thinning. *arXiv preprint arXiv:2110.01593*, 2021.

[5] Albert Gong, Kyuseong Choi, and Raaz Dwivedi. Supervised Kernel Thinning. *arXiv preprint arXiv:2410.13749*, 2024.

[6] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.

[7] Bharath K Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert RG Lanckriet. Hilbert space embeddings and metrics on probability measures. *The Journal of Machine Learning Research*, 11:1517–1561, 2010.

[8] Blanca Cano, Ángela Fernández, and José R Dorronsoro. Analysis of Kernel Thinning for scalable Support Vector Machines. In *International Conference on Hybrid Artificial Intelligence Systems*. Springer, 2025.